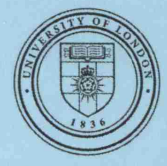


UNIVERSITY OF LONDON THESIS

Degree PhD Year 2007 Name of Author SMITH, Colette Joanne

Methodological issues when investigating the prevalence and incidence of antiretroviral-related toxicities amongst HIV positive individuals

THESIS
presented for the
DEGREE
of
DOCTOR OF PHILOSOPHY
in the Faculty of Medicine
(Field of study - Epidemiology)

COLETTE JOANNE SMITH

Royal Free Centre for HIV Medicine and
Department of Primary Care & Population Sciences,
Royal Free and University College Medical School, UCL
University of London

May 2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Declaration

I, Colette Joanne Smith, confirm that the work presented in this thesis is my own. Where information is derived from other sources, I confirm that this has been indicated in the thesis

Abstract

The introduction of highly active antiretroviral therapy (HAART) for the treatment of HIV infection has led to dramatic reductions in morbidity. However, these highly potent drugs have been associated with a number of side effects. It is important to accurately quantify the prevalence and incidence of these toxicities, and to be able to compare the impact of specific antiretrovirals on toxicity rates in an unbiased manner.

As we have become aware of potential HAART-related toxicities, there has been an increase in the frequency of monitoring of associated laboratory markers. Thus, diagnoses are made more quickly, and randomly abnormal values are more likely to be observed. There have also been changes over time in the specific antiretrovirals prescribed in HAART regimens. Consequently, newer antiretrovirals may seem to be associated with greater toxicity, even if this is not the case. Data simulations performed suggest that, when comparing the impact of specific HAART regimens on the occurrence of toxicity, considering the first measurement in a specified window leads to the most unbiased results.

The choice of most appropriate cut-off of a surrogate laboratory marker of toxicity can also lead to differences in prevlance estimates. My investigations suggest that the most appropriate cut-off for each toxicity and each laboratory marker must be considered individually, but that definitions in which a confirmatory measurement is required are often appropriate.

Results of a simulation study investigating a method proposed to account for unmeasured confounding, sample selection models, found this method gives unbiased treatment estimates in the situations we investigated. However, this method can lead to a lack of precision, and requires identification of a variable that is associated with treatment allocation. Thus, the use of these models in real-life settings may be limited.

Mis-specification of non-linear associations between variables in regression models can lead to biased estimates. I have found that the use of multi-fractional polynomials to systematically consider the most appropriate relationship between variables is a method that is easy to apply and leads to plausible results.

Acknowledgements

I would like to thank my supervisors, Caroline Sabin and Mike Youle, for their support and encouragement. I would like to especially thank Caroline Sabin, Andrew Phillips, Fiona Lampe and Amanda Mocroft for their helpful comments and advice.

Thanks to all who work at the Royal Free Hospital who have provided data for analyses in Chapters 4 to 6, especially Margaret Johnson and all staff at the Ian Charleson Day Centre. Thanks also to Sarah Walker, Julie Lanigan, Margaret Rhoads, Hermione Lyall and Diana Gibb for providing the data for Chapter 8 and for their input into the analyses, and to Scott Chavers, Jamie Robinson and Marianne Cunnington for their help with Chapter 7.

Thank you to the HIV Epidemiology and Biostatistics Unit at the Royal Free and University College Medical School, London, for their support and encouraging me not to give up: Sapna Shah, Wendy Bannister, Zoe Fox, Jani Purediredja, Alessandro Cozzi Lepri, Loveleen Bansi, Elaine Harris, Andy Thomson and Hannah Green.

The work carried out in Chapter 6 was supported by an unrestricted grant from Bristol Myers Squibb, and the work carried out in Chapter 7 was supported by an unrestricted educational grant from Glaxo Smith Kline. Other chapters were supported in part by the St Stephen's AIDS trust.

Last, but not least, thanks to my family and friends, especially mum, dad, grandma, Teresa, Paul Chapman, Marie and Paul. Thanks for all your support throughout the years I've spent studying, and thanks especially for not complaining when I've been late home from work recently – I got it finished eventually.

Chapters

Contents

Tables and Figures

Chapter 1 – Introduction

1.1 The natural history of Acquired Immunodeficiency Syndrome and Human Immunodeficiency Virus

1.1.1 Acquired immunodeficiency syndrome (AIDS) and human immunodeficiency virus (HIV)

Since acquired immunodeficiency syndrome (AIDS) was first observed in 1981[1] our knowledge of this disease has increased dramatically. AIDS occurs when an individual's immune system has been weakened sufficiently so that opportunistic infections (OIs) and lymphomas are able to develop.

The retrovirus responsible for AIDS, human immunodeficiency virus (HIV), was first identified in 1983 [2]. The HIV virus spreads between individuals via the exchange of bodily fluids, and thus HIV-positive individuals are commonly infected through homosexual and heterosexual sexual intercourse, by mother to child (vertical) transmission, through intravenous drug use using contaminated needles, or following the receipt of contaminated blood products.

It has been reported that around 30% of individuals experience some kind of illness when first infected with HIV. Individuals typically experience 'flu-like' symptoms, such as fever, malaise, night sweats and generalised lymphadenopathy [3]. This acute phase of infection is followed by a chronic phase. At first, individuals are asymptomatic (for around 10-12 years on average [4], although this varies greatly between individuals [5]), as the body manages to compensate for the depletion of the immune system. However, as the immune system weakens, opportunistic infections and lymphomas develop. An HIV-positive person is considered to have AIDS when one of a number of diseases that have been classified as AIDS-defining has occurred. The 1993 European AIDS surveillance definition is described in Figure 1.1. This definition is identical to that used in the United States of America, with the omission of the criterion which defines AIDS to have occurred when a CD4 cell count measurement of less than 200 cells/mm^3 has been recorded. As the immune system becomes even weaker, more opportunistic and AIDS-defining illnesses occur which can ultimately lead to death.

Figure 1.1 – 1993 European AIDS surveillance Case Definition [6]

1993 European AIDS Surveillance Case Definition for adults and adolescents aged greater than 13 years of age:
List of Indicator Diseases

- Bacterial infections, multiple or recurrent in a child under 13 years of age
- Candidiasis of bronchi, trachea, or lungs
- Candidiasis, oesophageal
- Cervical cancer, invasive
- Coccidioidomycosis, disseminated or extrapulmonary
- Cryptococcosis, extrapulmonary
- Cryptosporidiosis, intestinal with diarrhoea (> 1 month's duration)
- Cytomeglovirus disease (other than liver, spleen, or nodes) in a patient over one month of age
- Cytomeglovirus retinitis (with loss of vision)
- Encephalopathy, HIV-related
- Herpes simplex: chronic ulcer(s) (>1 month's duration); or bronchitis, pneumonitis, or oesophagitis in a patient over one month of age
- Histoplasmosis, disseminated or extrapulmonary
- Isosporiasis, intestinal with diarrhoea (>1 month's duration)
- Kaposi's sarcoma
- Lymphoid interstitial pneumonia in a child under 13 years of age
- Lymphoma, Burkitt's (or equivalent term)
- Lymphoma, immunoblastic (or equivalent term)
- Lymphoma, primary, of brain
- *Mycobacterium avium* complex, ir *M. kansasii*, disseminated or extrapulmonary
- *Mycobacterium tuberculosis*, pulmonary in an adult or adolescent (≥13 years)
- *Mycobacterium tuberculosis*, extrapulmonary
- Mycobacterium, other species or unidentified species, disseminated or extrapulmonary
- *Pneumocystis carinii* pneumonia (PCP) (also known as *Pneumocystis jiroveci*)
- Pneumonia, recurrent
- Progressive multifocal leukoencephalopathy
- *Salmonella* (non typhoid) septicaemia, recurrent
- Toxoplasmosis of brain in a patient over one month of age
- Wasting syndrome due to HIV

1.1.2 The life cycle of the HIV virus

The life cycle of the HIV virus is described in Figure 1.2. The HIV virus acts by infecting cells within the body, primarily the CD4 T-lymphocyte cells, which are involved in co-ordinating the body's immune response to antigens. The HIV virion contains two ribonucleic acid (RNA) copies of its genome within the virus particle. The virus enters the CD4 cell by attaching to its receptors and co-receptors (particularly the CCR5 and CXCR4 co-receptors). Reverse transcriptase is used to make a deoxyribonucleic acid (DNA) copy of the viral RNA genomes within the infected cell, which enables the HIV to be integrated into the DNA in the nucleus of the cell; a process enabled by the presence of integrase. The strands of the viral DNA in the nucleus separate and messenger RNA (mRNA), polypeptides and protease are created. The protease cleaves polypeptides into functional HIV proteins to create new HIV virions, which then bud from the cell surface and can go on to infect other cells. In this process of replication the CD4 cell subsequently dies, although the mechanism that causes this is still poorly understood. Thus, HIV mediates the destruction and depletion of CD4 cells in the body. Eventually, the production of compensatory T-cells fails and so the immune system collapses [7].

Figure 1.2 – The replication cycle of HIV; from Weiss et al [8]

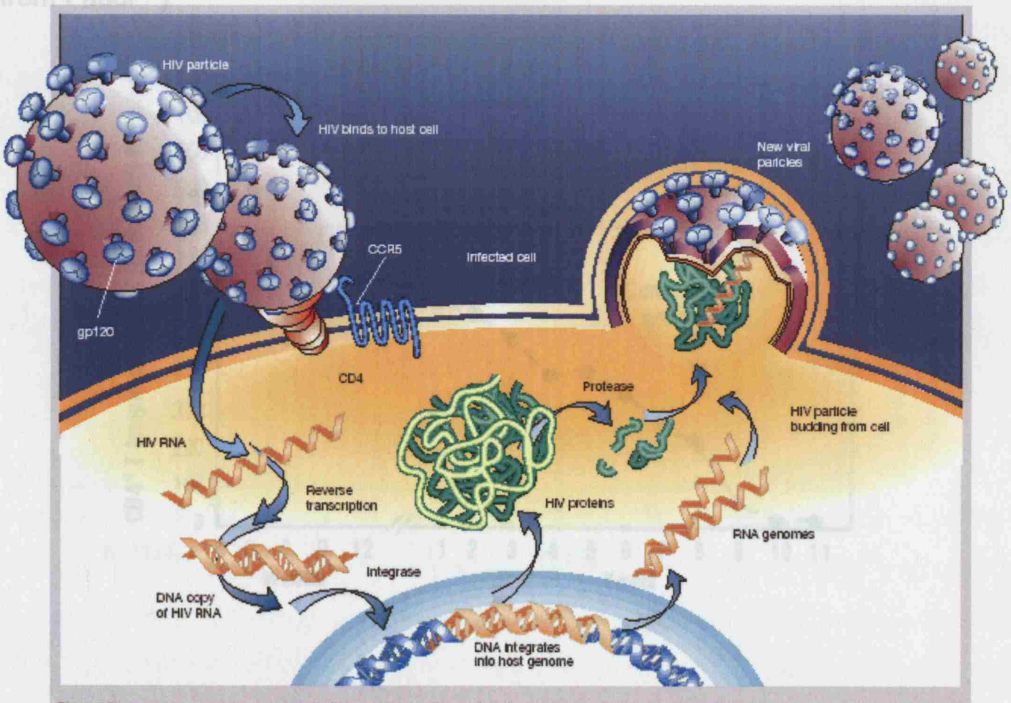
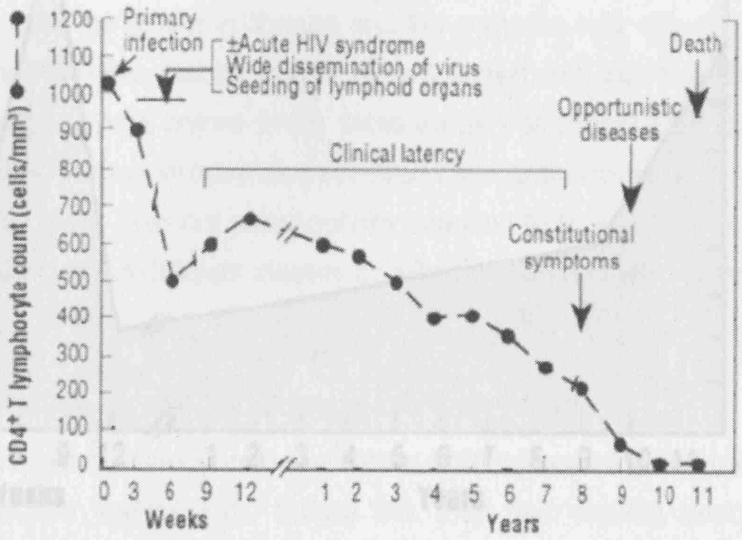


Figure 1 The human immunodeficiency virus. Replication cycle of HIV showing multiple steps available for attack by antiretroviral drugs.

1.1.3 Surrogate markers of clinical progression; the CD4 cell count and the HIV-RNA viral load

Although it is extremely variable, the number of CD4 cells present in HIV-negative individuals is in the range of 600-1500 per mm^3 [9;10]. It has been shown that HIV-positive individuals have much lower CD4 cell counts than HIV-negative individuals, and several cohort studies of HIV infected individuals before the introduction of antiretroviral treatment have shown that CD4 lymphocyte cell numbers gradually decline during HIV infection [10-24]. Furthermore, the CD4 cell count has been shown to be a good surrogate marker for clinical endpoints in HIV positive individuals, with those with CD4 cell counts of less than 200 cells/mm^3 being at a particularly high risk of developing AIDS [9;10;22;25]. Figure 1.3 illustrates the natural history of the CD4 count from the time of infection until the time of death in the absence of treatment. The CD4 cell count is only one marker of immunological function, and others, such as the CD8 cell count, the subsets of naïve and memory CD4 cells and other activation markers, have also been shown to be good markers of the immunological status of HIV-positive individuals [26-30].

Figure 1.3 – The natural history of the CD4 cell count during HIV infection (Adapted from Fauci [31])

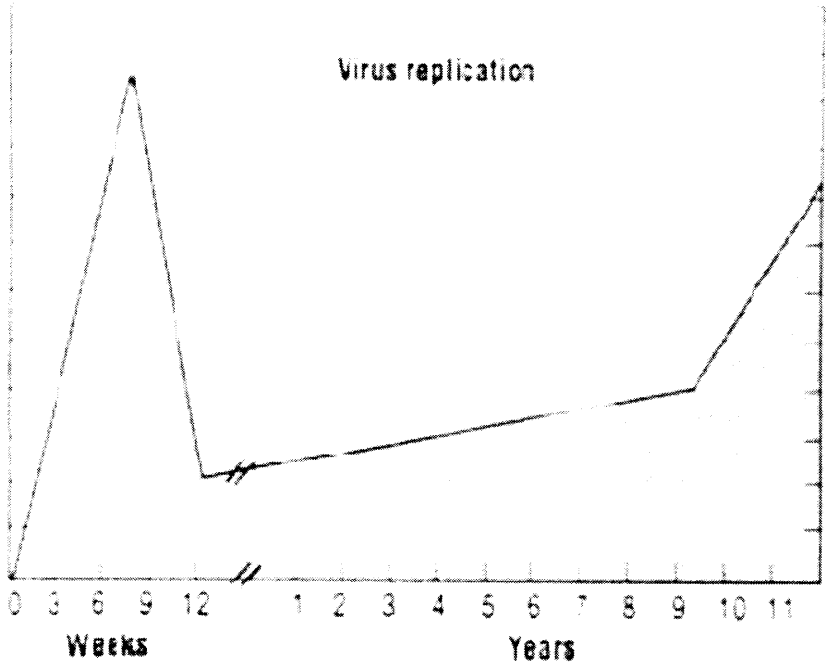


Assays that enable the measurement of the amount of HIV RNA copies in the plasma (viral load) were introduced in 1996. It has been established that during primary

infection the viral load can be as high as 10 million copies/ml, before it falls at the end of the acute phase following primary infection [31]. During the chronic phase of infection, there is a gradual increase in the level of HIV RNA present in the plasma. There is a rapid turnover of new virions being created, estimated to be in the region of 1 to 10 billion copies of virus each day [32-34]. The natural history of the viral load during HIV infection in the absence of treatment is illustrated in Figure 1.4.

Studies have shown that, independently of the CD4 cell count, the viral load is a good predictor of clinical disease progression [35-37], with those with higher viral loads generally having faster rates of disease progression to AIDS and death [38]. Initially, assays were only able to quantitate levels of HIV RNA in the plasma when they were above 400 or 500 copies/ml. However, assays with a lower limit of detection of 50 copies/ml were introduced in 1999, and, more recently, assays with lower limits of detection of 10 or even 3 copies/ml have been introduced.

Figure 1.4 – The natural history of the HIV RNA viral load during HIV infection (from Fauci [31])



1.1.4 Other factors associated with clinical progression

Older age has also been shown to associated with greater rates of clinical progression [39-42] (although this has not been corroborated in all studies [43-45]). Additionally, various studies have found associations with other factors, including co infection with cytomegalovirus [39], low concentrations of albumin [46], low concentrations of haemoglobin [47-50], female gender [51-55] and black and Hispanic ethnicity [54;55] (although this may be related to access-to-care [56]) all of which are associated with greater rates of progression to HIV and AIDS. Furthermore, those individuals who possess a 32-bp deletion in the CCR5 coding (the CCR5 Δ32 mutation), one of the co-receptors used by the HIV virion for cell entry, appear to have a better outcome. Those who are homozygous for CCR5 Δ32 are strongly, although not entirely, protected from HIV infection [57-60], and those who are heterozygous appear in some studies [61-64], but not all [65-67], to have slower rates of progression by about two years.

1.2 Antiretroviral therapy and highly active antiretroviral therapy (HAART)

HIV-positive individuals are treated with antiretroviral drugs (ARVs) in order to combat HIV infection. These ARVs act in different ways to inhibit replication of the HIV virus. There are three main classes, although others from other classes are gradually being introduced. Table 1.1 shows the licensed ARVs in the United Kingdom, the date they were licensed, the common dosage and the common side effects. Although the dates shown are the dates that the ARVs were licensed, HIV-positive individuals may have received these drugs before these dates as part of clinical trials or expanded access programmes. They may have also received other antiretroviral agents as part of clinical trials, which were then not subsequently licensed to treat HIV in the UK. This section briefly describes the different classes of ARVs drugs and their mechanism of action.

1.2.1 Nucleoside reverse transcriptase inhibitors (NRTIs)

Nucleoside reverse transcriptase inhibitors (NRTIs) were the first class of ARVs shown to be effective against HIV. Before this they had already been licensed to treat malignancies and herpes virus infections. NRTIs contain nucleotides, which are used when the viral HIV RNA is converted into DNA inside the CD4 cell. When the reverse transcriptase uses the nucleotides present in the NRTIs to transcribe the RNA rather than the viral nucleotides, the new DNA cannot be built correctly, and so the HIV virion is unable to replicate. The NRTIs that are currently licensed in the United Kingdom are

zidovudine (AZT), lamivudine (3TC), stavudine (d4T), didanosine (ddI), zalcitabine (ddC), abacavir (ABA) and emtricitabine (FTC). Combinations of NRTIs in a single pill have also been licensed, thus reducing pill burden. The combinations currently licensed are combivir (CBV; AZT with 3TC), trizivir (TZV; AZT+3TC+ABA) and kivexa (ABC+3TC).

1.2.2 Protease inhibitors

Protease inhibitors (PIs) act to inhibit the replication of the HIV virus at a different point in the life cycle to NRTIs. As the name suggests, they do so by inhibiting the function of the protease enzyme. Thus, although new HIV virions are created, they are unable to go on to infect other cells. The PIs currently licensed in the UK are indinavir (IDV), saquinavir (SQV) invirase (hard gel capsules), SQV fortovase (soft gel capsules), ritonavir (RTV), nelfinavir (NFV), amprenavir (APV), atazanavir (ATA), tipranavir (TPV), fos-amprenavir (FPV) and darunavir (DRV). Saquinavir invirase has been shown to be less potent than other PIs, as it only has approximately 4% bioavailability [68;69]. There is also evidence that nelfinavir may be less efficacious than other PIs [70;71]. A further PI, Kaletra has also been licensed, which consists of a single pill containing the PI lopinavir (LPV) with a small dose of ritonavir, which acts to boost and stabilise lopinavir levels in the blood. Ritonavir has also been shown to boost and stabilise levels of other PIs in the blood [72], and so RTV-boosted regimens, containing a PI taken with a small amount of RTV, have been introduced, and are now included in the British HIV Association's guidelines as appropriate regimens to be taken as part of first-line therapy [73].

Table 1.1 – Antiretrovirals licensed in the United Kingdom

Antiretroviral agent	Brand name	Abbreviation	Date UK license	Dose	Common side effects
Nucleoside reverse transcriptase inhibitors (NRTIs)					
Zidovudine	Retrovir©	AZT or ZDV	1986	300 mg BID	Nausea, headache, fatigue, anaemia, neutropenia, neuropathy
Didanosine	Videx EC©	ddI	1991	400 mg OD on empty stomach (>60 kg body weight)	Peripheral neuropathy, rare pancreatitis, avoid alcohol
Zalcitabine	Hivid©	ddC	1992	0.375-0.75 mg TD	Peripheral neuropathy, rarely used due to toxicity
Stavudine	Zerit©	D4T	1994	40 mg BID (>60 kg body weight)	Peripheral neuropathy (1-4 in early studies, 24% in expanded access)
Lamivudine	Epivir©	3TC	1995	150 mg BID	Generally well tolerated
Abacavir	Ziagen™	ABC	1998	300 mg BID	Hypersensitivity reaction, fever, malaise, possible rash, GI, respiratory
Emtricitabine	Emtriva	FTC			
Protease Inhibitors (PIs)					
Saquinavir hard gel	Invirase©	SQV (HGC)	1995	Used with ritonavir	Poor absorption, so rarely used
Indinavir	Crixivan©	IDV	1996	800 mg TID on empty stomach or with light snack	Kidney stones in 6-8%, occasional nausea, GI upset
Ritonavir	Norvir©	RTV	1996	600 mg BD with fatty food	Nausea, diarrhoea, numb lips, occasional hepatitis, hyperlipidaemia
Saquinavir soft gel	Fortovase©	SQV (SGC)	1997	1600 mg BD or 1200 mg TD	
Nelfinavir	Viracept©	NFV	1997	1250 mg BID or 750 mg TD with food	Diarrhoea, nausea
Amprenavir	Agenerase™	APV	1999	1200 mg BID	Rash, diarrhoea, nausea
Lopinavir+ritonavir	Kaletra©	LPV	2000	400mg LPV+100mg RTV BID with food	GI side effects common but mild, hyperlipidaemia
Tipranavir	Aptivus™	TPV	2005	500 mg+200 mg RTC BID with food	Raised lipids, liver disease
Atazanavir	Reyatz®	ATV	2003	400 mg OD or 300 MG+100 mg RTV OD with food	High levels of bilirubin,
Fos-amprenavir	Telzir (Europe) Lexiva™ (USA)	FPV	2003	1400 mg BID or 700 mg+100 mg RTV BID or 1400 mg +200 mg RTV OD	Raised triglycerides,
Darunavir	Prezista	DRV	2007	600 mg BID taken with 100mg RTV bid with food	Diarrhoea, nausea, headache, skin rashes
Non-nucleoside reverse transcriptase inhibitors (NNRTIs)					
Nevirapine	Viramune©	NVP	1996	200 mg BID or 400 mg OD	Transient rash, hepatitis
Efavirenz	Sustiva™	EFV	1998	600 mg OD, initially at bedtime	Initial dizziness, insomnia, transient rash
Nucleotide reverse transcriptase inhibitors (NtRTIs)					
Tenofovir disoproxil fumarate	Viread™	TDF	2001	300 mg OD with food	Generally well tolerated
Fusion (Entry) inhibitors					
Enfuvirtide	Fuzeon®	T-20	2003	90 mg BID Sub-cutaneously	Skin reactions at injection site

OD=once daily; BID=twice daily; TD=three times daily

1.2.3 Non-nucleoside reverse transcriptase inhibitors

Non-nucleoside reverse transcriptase inhibitors (NNRTIs) act by attaching themselves to the hydrophobic pocket of reverse transcriptase, and so they prevent the enzyme from converting RNA to DNA. As a result, the viral DNA cannot be incorporated into the cellular DNA, and so the cell is prevented from producing new virus. There are currently only two licensed NNRTIs in the UK; nevirapine (NVP) and efavirenz (EFV). Although two further NNRTIs, delavardine (DLV) and loviride (LOV), have been used to treat some patients in the UK, these antiretrovirals were not given a European license. Additionally, etravirine (TMC125) is in advanced development for treatment experienced patients.

1.2.4 Other classes of antiretroviral drugs

Whilst the most common antiretrovirals used to treat HIV-positive patients belong to the NRTI, PI and NNRTI drug classes, there are other antiretrovirals of different drug classes that are licensed to be used against HIV. One such class is nucleotide reverse transcriptase inhibitors (NtRTIs). Currently, the only NtRTI licensed in the UK is tenofovir disfumarate (TDF). NtRTIs act very similarly to NRTIs, except for the fact that NRTIs must undergo phosphorylation in order to become active in the body, whereas NtRTIs do not as they are already chemically active. TDF is also available as a single tablet with the NRTI FTC, which is known as Truvada.

Other classes of antiretrovirals include fusion or entry inhibitors (including the licensed drug gp120 inhibitor, enfuvirtide [T20; EFV]). Entry inhibitors prevent the HIV virion from attaching to the co-receptors of the CD4 cell, and thus the virus cannot enter the cell and replicate. T20 must be administered sub-cutaneously via an injection, and is currently only licensed for those with limited treatment options. Other entry inhibitors, such as the CCR5 inhibitor maraviroc [74;75] are also at an advanced stage of development. Finally, research into the feasibility of integrase inhibitors is ongoing, although there are currently no licensed drugs available in this class. However, clinical trials of these drugs indicate they are likely to be highly effective [76-78].

1.2.5 Treatment strategies: highly active antiretroviral therapy (HAART)

The first class of antiretrovirals to be introduced was the NRTIs, the first of which, AZT, was introduced in 1986. Initially, AZT was administered as monotherapy. The

Concorde trial in 1986 [79] in symptomatic HIV-positive patients found that AZT had a limited beneficial effect on overall survival in these patients. This was followed by trials investigating combinations of two NRTIs as dual therapy, including the Delta [80] and ACTG 175 trials [81], which showed that dual therapy led to greater survival benefits than monotherapy [80;82-84]. However, it was not until the introduction of combinations of antiretrovirals including one PI and two NRTIs (first studied in the ACTG 320 clinical trial which compared AZT+3TC with AZT+3TC+IDV [85]) in 1996 that substantial reductions in morbidity and mortality were seen [86-89]. Other regimens containing at least two NRTIs with either an NNRTI, the NRTI abacavir (although there is some evidence of greater treatment failure on this regimen [70;71]) and PIs boosted with RTV have also been shown to be highly efficacious. These regimens are commonly known as highly active antiretroviral therapy (HAART) or combination antiretroviral therapy (cART) regimens.

1.2.6 Surrogate markers for response to HAART

As clinical events in the HAART era are fortunately now rare, it is difficult to assess the efficacy of new antiretroviral drugs and drug regimen strategies using clinical endpoints. Therefore, the efficacy of antiretrovirals is now usually assessed using surrogate endpoints. In particular, the CD4 cell count has been shown to be a good prognostic marker of clinical status amongst those receiving HAART, similarly to its role as a prognostic marker in antiretroviral-naïve patients [90-95]. Individuals who maintain suppression of viral replication (attain viral loads of less than 400 or 50 copies/ml) have been shown to achieve the greatest CD4 cell increases, and this has been shown to be the strongest factor associated with better CD4 cell responses [91;96-100] and a better clinical response to HAART [91-93;101;102]. Indeed, it is generally accepted that the primary aim of HAART is to suppress the viral load to below the lower limit of detection, and thus the main endpoints of randomised controlled trials investigating treatment efficacy are usually the ability to achieve and maintain an 'undetectable' viral load.

1.2.7 Predictors of virological response to HAART

Several studies have investigated the factors associated with a good virological response to HAART. Response to the first HAART regimen received has been shown to result in the best chance of virological suppression, and as the number of previous antiretrovirals ever received increases, so the likelihood of achieving an undetectable

viral load decreases [103]. The reasons for this are primarily due to the emergence of drug-resistant virus, which is discussed in detail in sub-section 1.3.1.

One of the main predictors of a good virological response to HAART is a high level of adherence to the HAART regimen [104-109]. Individuals who contracted HIV via intravenous drug use [70;110;111], those with cognitive dysfunction [111;112], females [112-114] (although not in all studies [109]), younger patients [111] (although not in all studies [109]), and individuals of black African ethnicity [70;115;116] have been shown to be more likely to have lower adherence and thus subsequently experience a worse virological response. It has been estimated that maintaining adherence rates of around 95% is needed to ensure a good virological response to HAART [117].

The CD4 cell count at the time of starting HAART, or the CD4 nadir (the lowest CD4 cell count ever measured prior to starting HAART) have been shown to be associated with response to treatment. Those who start treatment with a CD4 cell count of less than 200 cells/mm^3 have been shown to be at a particularly high risk of experiencing clinical progression [103;103;118-122]. Some studies have also shown that higher pre-treatment viral loads [103;119-121], younger age [105;119] (although not all studies have found this association [103]), depression [105], being male [103], pre-treatment with mono- or dual-therapy [103;120;121;123] and non-white ethnicity [119] are associated with an increased risk of virological failure.

1.3 The limitations of HAART

Although the effects of HAART on survival have been dramatic, antiretrovirals have been unable to completely eradicate HIV infection. This is due both to the continuing production of low-level viraemia even amongst those achieving good virological responses [124-127] and to the presence of a subset of latently-infected CD4 cells which have the ability to lie dormant for several years before reactivating [128-130]. Therefore, as there is still no cure for HIV, it appears that HIV-positive individuals may have to remain on HAART regimens for the foreseeable future. Although recent studies have shown that individuals who maintain virological suppression experience a good immunological response to HAART up to four years and more after starting HAART [131;132], there appear to be two major threats to the ability to successfully remain on HAART regimens for prolonged periods of time. The first is that drug-resistant virus can develop over time, and the second is the toxicity and inconvenience of antiretroviral regimens.

1.3.1 Antiretroviral drug resistance

When transcribing the RNA genome into the viral DNA, the HIV virus is quite error prone and, as so many replications of the HIV virus are made daily, many spontaneous genetic mutations in the reverse transcriptase and protease regions of the virus can occur [133]. In the absence of treatment, one genetic form of the virus (the one that is able to replicate the most quickly), known as "wild type" virus, usually becomes the most prevalent. However, in an individual receiving antiretrovirals, if incomplete viral suppression is present (e.g. due to incomplete adherence, or a sub-optimal regimen) then genetic mutations that allow the virus to grow despite the presence of antiretrovirals may make that mutant strain the most prevalent. As a result, the individual's viral load will increase, as the antiretroviral will no longer be able to suppress viral replication. Mutations that favour the replication of HIV in the presence of a specific drug are called that drug's primary resistance mutations. For example, a mutation of the reverse transcriptase gene at the 184 position from an 'M' in the wild type virus to a 'V' (the M184V mutation) confers resistance to the NRTI lamivudine. Once antiretrovirals are discontinued in an individual with resistance mutations, the wild type virus again becomes the fittest virus and thus the most prevalent. However, the resistant virus remains a minority sub-species, and so if the antiretroviral is re-introduced the resistant virus will reoccur and complete virological suppression is unlikely to be possible [133]. Although there are over 15 antiretrovirals available, there are currently only three main classes, and virus that is resistant to one drug can be resistant to other drugs of the same class [134].

Antiretroviral resistance is most prevalent amongst those individuals who were exposed to mono or dual therapy before starting HAART, as these initial regimens were sub-optimal and thus full suppression of the viral load was usually not achieved [101]. However, drug resistance also occurs in those who received HAART as their first-line antiretroviral treatment. A recent study by the UK Collaborative Group on HIV drug resistance [135] found that, after 6 years of HAART amongst previously antiretroviral-naïve individuals, 24% had at least one mutation that would lead to resistance to an NRTI, 10% possessed at least one PI mutation, 16% possessed at least one NNRTI mutation and 20% had detectable mutations to two of the three main drug classes. Furthermore, drug-resistant virus can be transmitted to individuals when they first become infected with HIV, before ever being exposed to antiretrovirals [136].

1.3.2 Antiretroviral-related toxicities

Although ARVs have led to dramatic decreases in HIV-related morbidity and mortality, they have also been associated with a number of toxicities. Toxicities can lead to lower rates of adherence [137], which can in turn lead to incomplete virological suppression, development of drug resistance, and virological and treatment failure. They can generally be described as short-term or longer-term toxicities and can range from mild to potentially life threatening.

Some short-term toxicities have been reported to be associated with specific antiretrovirals or antiretroviral drug classes (Table 1.1). Around 8 per cent of individuals receiving the NRTI abacavir experience a hypersensitivity reaction and rash which is potentially fatal, and does not resolve on re-introduction of the antiretroviral [138;139]. Efavirenz has been associated with a risk of dysphoria (vivid dreams, nightmares, mood disturbance, drowsiness and disorientation) [140]. Nevirapine has been associated with rash and a risk of hepatitis, especially amongst women with CD4 cell counts of more than 250 cells/mm^3 [141]. Kaletra has been associated with gastrointestinal side effects [73]. Atazanavir has been associated with hyperbilirubinaemia and jaundice [73;142]. NRTIs in general, and particularly the so-called "d-drugs" d4T, ddI and ddC, have been associated with mitochondrial toxicity and peripheral neuropathy [143]. ZDV has been associated with anaemia [144]. T20 has been associated with injection site reactions as it must be administered sub-cutaneously [145;146]. Milder toxicities, such as nausea, vomiting, diarrhoea, anaemia, headaches and fatigue have been associated with antiretrovirals in all three main drug classes [73].

Longer-term toxicities have also been associated with individual antiretrovirals, specific antiretroviral drug classes and antiretroviral treatment in general. These include hepatotoxicity, cardiovascular disease, lipodystrophy and renal disease and are reviewed in more detail in chapter 2.

1.4 HIV infection in children

1.4.1 The transmission of HIV to infants and children, and the natural history of HIV disease

The suspected route of transmission for the majority of children who are HIV positive is via mother-to-child transmission. It is estimated that around 15%-30% of babies born to HIV-positive mothers will become infected with HIV if the mother does not receive antiretroviral treatment during pregnancy, and a further 10-20% will become infected

through breast-feeding [147]. The risk of acquiring HIV infection can be greatly reduced to less than 2% by antiretroviral treatment for the mother during pregnancy, elective caesarean section, prophylactic treatment for the newly-born infant, and by bottle-feeding [148;149].

The natural history of HIV in children is different to that in adults. Great variability in the rate of disease progression has been observed [150-153], but children generally progress to AIDS more quickly than adults do, as their immune system is still developing [73]. It has been estimated that between 47% [150] to 64% [151] of HIV-positive children will remain AIDS-free at the age of 6 years. The definition for AIDS-defining illnesses in children are similar to those in adults, and are described in Figure 1.5 below.

Figure 1.5 – Summary of the European Case Definition for AIDS surveillance in Children; Revision 1995 [154]

European Case Definition for AIDS surveillance in Children;
List of indicator Diseases
Revision 1995

I. Without Laboratory Evidence for HIV infection and no other cause of immunodeficiency

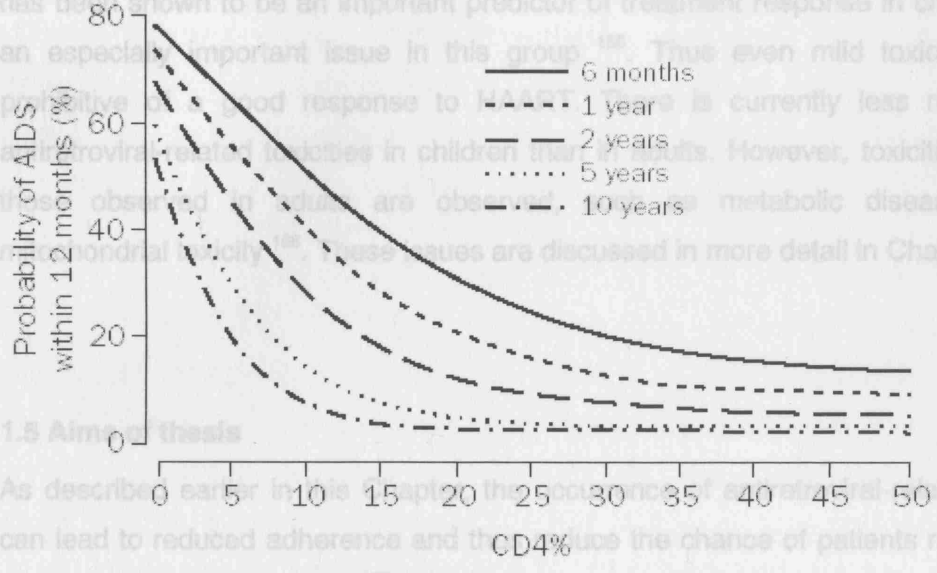
- Candidiasis of the oesophagus, trachea, bronchi or lungs
- Cryptococcosis, extrapulmonary
- Cryptosporidiosis with diarrhoea persisting >1 month
- Cytomegalovirus disease of an organ other than liver, spleen, or lymph nodes on a child>1 month of age
- Herpes simplex virus infection causing a mucocutaneous ulcer that persists longer than 1 month; or bronchitis, pneumonitis, or oesophagitis for any duration affecting a child >1 month of age
- Kaposi's sarcoma
- Lymphoma of the brain (primary)
- Lymphoid interstitial pneumonia or pulmonary lymphoid hyperplasia (LIP/PLH complex)
- *Mycobacterium avium* complex or *M. kansasii* disease, disseminated (at a site other than or in addition to lungs, skin, or cervical or hilar lymph nodes)
- *Pneumocystis carinii* pneumonia
- Progressive multifocal leukoencephalopathy
- Toxoplasmosis of the brain affecting a child >1 month of age

II. With laboratory evidence of HIV infection

- Any disease listed above
- Serious bacterial infections, multiple or recurrent, of the following types: septicaemia, pneumonia, meningitis, bone or joint infection, or abscess of an internal organ or body cavity
- Coccidioidomycosis, disseminated (at a site other than or in addition to lungs, skin, or cervical or hilar lymph nodes)
- HIV encephalopathy
- Histoplasmosis, disseminated (at a site other than or in addition to lungs, skin, or cervical or hilar lymph nodes)
- Isosporiasis with diarrhoea persisting >1 month
- Lymphoma, small, noncleaved cell (Burkitt's) or immunoblastic or large cell lymphoma of B-cell of unknown immunologic phenotype
- Any mycobacterial disease cause by mycobacteria other than *Mycobacterium tuberculosis*, disseminated (at a site other than or in addition to lungs, skin, or cervical or hilar lymph nodes)
- *Mycobacterium tuberculosis*, disseminated or extrapulmonary
- *Salmonella* (nontyphoid) septicaemia, recurrent
- HIV wasting syndrome

The use of CD4 cell counts as a marker of immunological function and as a surrogate for HIV disease progression is more difficult amongst children than in adults, as CD4 cell counts vary according to age [10;155]. A study of HIV-negative children found the mean (standard deviation) CD4 cell count in children aged 1-2 years to be 2530 (648) cells/mm^3, and the mean number of CD4 cells decreased with increasing age [10]. Therefore, a different surrogate marker of clinical progression is required. The age-adjusted CD4 z-score may be used, which adjusts the CD4 value for the age of the child. A CD4 z-score of 0 implies that the child has the average CD4 cell count for their age; each increase or decrease of 1 corresponds to an increase or decrease of 1 standard deviation away from the average CD4 cell count for that age. Another measure of immunological function used is the CD4 percentage (as a percentage of the total number of T-lymphocytes), which has been shown to remain at a more constant level throughout childhood and into adulthood in HIV-negative individuals [10], and it is able to predict risk of progression to death in HIV-positive children [156]. However, it has also been shown that the risk of progression to AIDS and death according to the current CD4 cell percentage still varies according to age (Figure 1.6) [157].

Figure 1.6 – Probability of developing AIDS within the next 12 months by age and current CD4 percentage-from Dunn et al [157]



It has been reported that the HIV RNA viral load reaches a peak in HIV-positive children at between 100 000 to 1 000 000 copies/ml at two months of age. In contrast to adults who experience a rapid decline from the peak viral load level to the viral set point when infected with HIV, it takes around two years for the viral load to decline to

more 'stable' levels [158-160]. Similarly to adults, the HIV RNA viral load in the absence of treatment is additionally a good marker of clinical progression in HIV positive children in addition to the CD4 cell percentage, with higher levels being associated with an increased risk of disease progression [156;157;160-162].

1.4.2 HAART for HIV positive children

The antiretrovirals licensed to treat HIV-positive children are the same as those licensed to treat adults, albeit at modified doses. However, there are no adequately powered trials allowing direct comparison of different HAART regimens in children, and thus evidence of antiretroviral efficacy largely comes from observational data and clinical trials of adults [155]. Children have been shown to experience good increases in CD4 cell counts when receiving HAART and achieving a virological response [163-165], although not to the extent of that observed in adults [166].

1.4.3 Antiretroviral-related toxicities in children

The issues of antiretroviral-related toxicities are especially important for HIV-positive children, as they will potentially be exposed to HAART for their lifetimes. Adherence has been shown to be an important predictor of treatment response in children and is an especially important issue in this group [155]. Thus even mild toxicities can be prohibitive of a good response to HAART. There is currently less research into antiretroviral-related toxicities in children than in adults. However, toxicities similar to those observed in adults are observed, such as metabolic diseases [167] and mitochondrial toxicity [168]. These issues are discussed in more detail in Chapter 8.

1.5 Aims of thesis

As described earlier in this Chapter, the occurrence of antiretroviral-related toxicities can lead to reduced adherence and thus reduce the chance of patients maintaining a good response to HAART [137]. The toxicities themselves can also lead to reduced quality of life, and may even be life threatening [138;139;141]. Therefore, it is important to accurately quantify the risk of these adverse events, to compare the incidence of toxicities associated with different antiretrovirals, and to identify factors associated with an increased risk of toxicity. However, there are a number of methodological issues that can hinder these analyses. Thus, the aim of this thesis is to investigate the

potential biases that can occur when investigating antiretroviral-related toxicities and propose methods that can account for these biases.

I shall begin in Chapter 2 by considering the published literature and summarising the methodological issues that emerge, and thus generate hypotheses surrounding important potential biases which will be investigated further. Next in Chapter 3, I shall describe the datasets upon which analyses in this thesis are performed. In Chapter 4, I shall consider changes over time in terms of the frequency of monitoring of toxicities and the impact of this on any analyses. In Chapter 5 I shall investigate definitions of outcome measures and their impact on the observed prevalence and incidence of antiretroviral-related toxicities. Chapter 6 considers the association between demographic and treatment factors with the occurrence of antiretroviral-related toxicities. In Chapter 7 I shall describe Sample Selection models, a proposed method to account for unmeasured confounding in observational datasets. For Chapters 4 to 7, I shall focus on previously antiretroviral-naive HIV-positive adults starting HAART for the first time, and toxicities occurring in the first year of HAART. However, in Chapter 8 I will study the use of fractional polynomials for accounting for non-linear associations between variables, focusing on an example of the impact of age on the occurrence of antiretroviral-related total cholesterol changes in HIV-positive children. Finally, in Chapter 9 I will summarise the findings of this thesis.

Chapter 2 – The problems faced when assessing the prevalence and incidence of antiretroviral-related toxicities

2.1 Introduction

As explained in Chapter 1, the widespread use of HAART has transformed the treatment of HIV and AIDS, and has resulted in dramatic reductions in mortality and morbidity [87]. HAART has, to a great extent, led to a switch in the way that HIV is managed, from being treated as a terminal disease to a chronic disease [86;89;169]. However, although the benefits of HAART cannot be disputed, these highly potent agents have been associated with many severe and possibly life threatening toxicities. These toxicities have become more apparent as patients remain on HAART for longer periods of time. Possible side effects that have been identified include fat redistribution, insulin resistance, dyslipidaemia, lactic acidosis, hepatotoxic effects, renal disease and bone disease [170-176].

HAART-related toxicities have been investigated in a number of studies. There has been great variability between these studies in the reported incidence and prevalence rates of these toxicities, even amongst those on similar drug combinations. This chapter explores possible explanations for these differences. As there are several different toxicities, I have chosen to concentrate on three of the most common. Therefore, I have examined possible biases and differences between studies that could explain the variations observed in the reported incidence of HAART-associated metabolic, hepatotoxic, and renal disorders.

2.2 Literature search methods

The papers included in this review were identified by searching the Medline and Pub Med databases. The search criteria used were as broad as possible in an attempt to ensure that all relevant papers were identified. The search terms used were *HIV* and either *HAART* or *antiretroviral* along with any of the following: *side effects*, *toxicity*, *renal*, *cardiovascular*, *nephrology*, *hepatic*, *liver*, *heart*, *kidney*, and *cholesterol*. I identified the results of these searches and assessed the relevance of each study by reading the titles and on-line abstracts of the papers. Further possible papers were found by additional hand searching of *AIDS*, *JAIDS*, *New England Journal of Medicine*, *Lancet*, *Journal of Infectious Diseases*, *Antiviral Therapy*, *JAMA*, and *HIV Medicine*. The main points of the relevant papers identified in the search were then summarised.

In this review I explore the features of studies that could lead to differences in observed incidence and prevalence rates of HAART-related toxicities and then discuss the merits and limitations of the different approaches. The features generally fall into two categories: general issues that applied to all studies investigating HAART-related toxicities, and factors that were specific to the disorder being studied. I start by including a section in which I explore general methodological issues, followed by specific sections relating to each specific disorder.

2.3 Methodological Issues: Study Design: Randomised controlled trials versus observational studies

2.3.1 Randomised controlled trials

It is well known that randomised controlled trials (RCTs) are the "gold standard" for assessing causality [177-179]. The main reason for this is that randomisation ensures that any differences observed between the two (or more) arms of a trial at baseline occurred solely by chance. Therefore, provided the care given to individuals in each arm during the trial is identical except for the intervention being studied, any differences in outcome can be attributed to the intervention itself. RCTs also have the benefits of being able to monitor individuals closely at regular time intervals, and to collect very detailed information on adverse events.

However, RCTs have some limitations. It is worth noting that some of these limitations may apply specifically to RCTs carried out in the HIV setting. Firstly, as they are costly and require intensive follow up, they tend to only observe patients for short periods of time. Therefore, meaningful information about clinical outcomes that can take decades to manifest, such as death, are rarely available. Further, if a patient with mild symptoms of a HAART-related disorder switches treatment regimens at an early stage before more serious signs have become apparent, it is unlikely that many trial analyses would capture this, particularly if follow-up ceases at this point. As a result, RCTs may underestimate the true incidence of toxicities in these patients.

Most RCTs are designed, and therefore powered, to evaluate the efficacy of the antiretrovirals being studied. The prevalence of toxicities and adverse events are usually included as secondary endpoints [180-183]. Therefore, RCTs are often not powered for detecting differences in incidence of toxicities, and so any estimates obtained may not be precise.

It is common in the HIV setting to carry out unblinded, or open, RCTs in which clinicians and patients are aware of the treatment arm the patient is in [184-186]. In many cases it may be difficult to blind patients and clinicians to treatment, as many antiretrovirals have specific side effects that immediately identify the randomised treatment. This lack of blinding may create a potential source of bias, as information collected on adverse events may be affected by preconceptions of the toxicities associated with each antiretroviral.

When designing an RCT to assess HAART-related toxicities the choice of an appropriate control arm is not immediately clear. The true question of interest when investigating the incidence of HAART-related toxicities is "is the incidence of toxicity higher amongst those on HAART than it would have been had they not received it", but very few studies include a control group who do not receive antiretroviral therapy for a long period of time, as their immunological and virological well-being would be compromised. RCTs can only compare the rates of disorders amongst those receiving different antiretroviral combinations, which may limit the conclusions of the trial. For example, if the incidence of hypercholesterolaemia is the same in two treatment arms of a trial, it may be difficult to establish whether this is because both treatments lead to similar high rates or because both treatments do not raise cholesterol. Identifying a suitable non-trial control group with similar characteristics may be problematic.

It is well known that a major limitation of RCTs lies in the type of patients who are recruited to trials. It has been shown that those excluded from RCTs are very different to those who are included [187-190]. For example, Madge et al [189] studied those who participated in HIV clinical studies in a single treatment centre (the Royal Free Hospital, London) over 12 years. Although they observed high rates of participation in clinical trials, it was noted that intravenous drug users (IDUs) were under-represented. Furthermore, all trials have inclusion and exclusion criteria, which mean that those who are at the most risk of adverse events may be excluded from participating in trials [185;191-194]. For example, the CNAB3005 entry criteria [193] required that included individuals' amylase levels were less than 1.5 times the upper limit of normal (ULN), that levels of hepatic aminotransferases were less than 5 times the ULN, neutrophil counts exceeded 1000/µl and that their estimated creatinine clearance was greater than 40 mL/min. Although differences in the study population will not lead to biased comparisons, they may lead to an underestimate of the incidence rates present in the general HIV positive population on HAART and may explain why incidence rates are different to those reported from observational studies.

2.3.2 Observational studies

Although RCTs will and should remain the gold standard for assessing causal relationships between HAART and toxicities, observational studies also play an important role. Observational studies, such as cohort studies and case-control studies, study individuals where interventions are administered according to patient and clinician wishes, such as the Royal Free Hospital Database. The main benefits of observational studies are that they are able to follow complete clinic populations for long periods of time. Therefore, all HIV positive patients should be represented. However, it must still be remembered that, for example, a large city-based clinic established especially for HIV positive individuals may have a very different demographic and behavioural profile to smaller rural-based clinics elsewhere in the world, and so studies may still vary in this respect. A further advantage is that, in general, observational studies are designed to have longer periods of follow-up, and so can look at the longer-term consequences of HAART, although patients will still switch regimens if their laboratory markers begin to suggest that a toxicity is likely to be occurring.

The major limitation of observational studies is that in most cohorts the reason for the choice of any particular regimen in any given individual is often unknown, creating a potential source for bias. For example, consider the case where there are two regimens which both lead to similar rates of myocardial infarction (MI), but where a clinician would assign patients with elevated total cholesterol levels to the first regimen as (s)he felt that the second regimen had a worse toxicity profile. In this situation, an investigator comparing the two regimens would be likely to find that those on the first regimen had a higher rate of MI than those on the second regimen. This would, of course, be a consequence of the fact that those on the first regimen had higher baseline total cholesterol levels and thus were already at higher risk of MI, even prior to treatment, rather than reflecting a direct influence of the first regimen. Although there are statistical methods that can account for confounders, such as total cholesterol in the example above, a confounder can only be adjusted for if it is known and has been measured. This issue is the focus of Chapter 7.

As for RCTs, it is difficult to choose an appropriate control group when examining HAART-related disorders in observational studies. In particular, if the question of interest is whether HAART "causes" toxicity-related disorders, then this may be hard to address in such studies. Calculating the rate of adverse events amongst those on

HAART compared to those who have not yet started antiretroviral therapy may not be appropriate as the groups may have very different demographic or clinical profiles, for example with regard to CD4 cell counts. The D:A:D study [195] and the SMART study [196] attempted to address this by computing risk ratios for cardiovascular events for each additional year of combination antiretroviral therapy received, and controlled for markers of infection (such as nadir CD4 cell count and time since diagnosis) as those with most extensive exposure to treatment would be expected to have more advanced disease. Therefore, HIV positive untreated patients in this study (either those who remained untreated or treated patients prior to treatment) served as controls. Another possibility to is compare an individual's levels of a particular laboratory marker (such as total cholesterol) before they started HAART to those whilst on HAART in order that each individual may act as their own control. However, this approach cannot examine whether a patient's markers would have changed anyway had the patient not required HAART at that particular time. It is also important to consider the fact that, as time of exposure to antiretrovirals increases, so other factors change at the same rate, such as a patient's age and calendar time. Extrapolating which of these factors are associated with changes in event rates is very difficult in observational studies as factors are highly correlated [197].

The amount of loss to follow-up and frequency of monitoring can vary greatly between studies. If loss to follow-up is not random, but occurs because a patient drops out of the study due to toxicity-related events without reporting this to their clinician, then this may lead to biased estimates of the incidence of toxicities. RCTs, in general, would have visits scheduled at least 4, 8, 12, 24 and 48 weeks after the trial began. Observational studies, however, will differ in when patients attend clinic according to several different factors including clinic policy, patient wishes and the perceived need for follow-up for that patient. For example, poorly adherent patients may be thought to benefit from more regular follow-up. A further issue is how closely disorders are followed. In RCTs all adverse events would be expected to be closely monitored as the trial protocol would specify the frequency of all laboratory monitoring and blood tests. However, in routine clinic practice this may depend on how important the treating physician or clinic perceives the risk of adverse events for that particular regimen. This may have an impact on the frequency of measurements and the likelihood of disorders being reported. Another limitation of observational data is that, in general, few cohorts collect detailed information on antiretroviral-related toxicities unless they are treatment limiting, although most collect laboratory data. Finally, adherence may not be as high in routine practice as it may be in a clinical trial, either because of the type of patients who are

enrolled or because trial procedures may encourage better adherence, and this may lead to lower rates of disorders being observed in observational studies in line with the lower rates of efficacy. These issues are discussed further in Chapter 4.

Whilst RCTs collect data prospectively by design, observational studies can collect data both retrospectively and prospectively, and this can impact on the incidence of HAART-related toxicities observed. If possible, it is preferable to collect data prospectively, as it is harder to control for potential biases in retrospective data. It is likely that retrospective data will contain more missing information, as the investigators cannot control or request the information that they want, and so have to rely on the information that was recorded at a time when the importance of toxicities may not have been fully understood, and when data on the particular toxicity may be unreliable. Data are also unlikely to be recorded in a uniform manner, and so there is the potential for the opinions of the investigator to influence the information collected. In a prospective study, the investigator can choose which information is relevant and ensure it is collected in a uniform manner, and thus the chance of differential ascertainment in different calendar periods is reduced.

In summary, the type of study used to assess the rates of HAART-related toxicities can impact on the results observed. The types of patients recruited to RCTs, and the short-term nature of such trials compared to the potential for unknown confounders in observational studies mean that both types of studies have benefits and limitations, and these must be remembered when considering the results of studies examining HAART-related toxicities.

2.4 Methodological Issues: Definition of outcome measures

It is possible to define toxicities in a number of ways. In general, there are two main ways in which they are defined: clinical endpoints and laboratory-defined toxicities. Although the papers that have looked directly at the occurrence of clinical events [195;198-201] perhaps have the most clinically useful outcome measure, these studies still suffer from limitations. They reflect the most extreme (but equally most sensitive and specific) outcome. However, clinical outcomes can take decades to manifest and so they tend to have low incidence rates. Therefore, large numbers and person-years of follow up are required for such a study to have sufficient power to detect a clinically relevant relationship with therapy. For example, the D:A:D study [195] calculated that, for their study to be sufficiently powered, information on 100 cardiovascular events in 30,000

person years of follow up was required. This, along with the fact that HAART was introduced relatively recently, means that few prospective studies with sufficient power to address these issues exist.

Therefore, a much more common approach to identifying potential toxicities is to consider changes in laboratory markers (such as total cholesterol, alanine aminotransferase [ALT] and creatinine levels) that have been shown, in HIV negative populations at least, to be associated with clinical outcomes [202-204]. Changes in these markers tend to occur over a much shorter time period than the occurrences of clinical events in those on HAART. However, there are differences in the ways that studies have identified and reported changes in laboratory markers. In particular, studies either report the number of individuals exhibiting laboratory markers above or below a pre-specified threshold [205-210], or report the average change from pre-therapy/baseline values [180;208;211-218]. Unfortunately, this has the effect of making comparisons between studies very difficult. For example, in one study [212] the median total cholesterol after 48 weeks of HAART fell from pre-therapy values by 0.89 (95% Confidence Interval [CI] -1.48 to –0.30) mmol/l whereas in another [210] after 24 weeks of HAART 4% of individuals had a total cholesterol level of greater than 6.2 mmol/l. It is not obvious how to compare these two results. Studies often also use different cut-offs to define an event (for example some studies consider hypercholesterolaemia to have occurred when the total cholesterol exceeds 6.2 mmol/l, whereas others consider a cut-off of 5.5 mmol/l). In addition, other studies have reported only that there were no significant differences in laboratory markers over time without reporting the average change itself [215;219;220]. This is investigated in more detail in Chapter 5.

It is possible that increases in laboratory markers are associated with a different increase in the risk of a clinical event in HIV positive individuals compared to HIV negative individuals. The relationship may also be different between HIV positive individuals receiving or not receiving HAART. To my knowledge, no study has assessed the relationship between clinical events and laboratory markers in HIV positive patients let alone in those on HAART. Thus, we have to make a major assumption that implications of raised (or lowered in the case of haemoglobin, albumin and others) laboratory markers have the same meaning in HIV positive individuals on therapy as they do in HIV negative individuals. If this is not true, then there is the risk that studies may over- or under- estimate the importance of HAART-related toxicities.

Although changes in laboratory markers tend to occur over a short time scale, not all changes will ultimately end in the occurrence of clinical events. For example, many

laboratory markers have poor specificity for clinical outcomes and some abnormal laboratory markers may resolve if left untreated. Therefore, any studies looking at changes in laboratory markers risk overestimating the importance of a particular toxicity. On the other hand, changes in laboratory markers may highlight the presence of HAART-related toxicities that are not as serious as a clinical event, but may still need addressing.

2.5 Methodological Issues: Changes over time

As HIV has come to be considered more as a chronic disease with a need for life-long treatment, so toxicities have taken on a greater importance. For example, the British HIV Association's (BHIVA) 2003 national clinical audit of treatment for people with HIV [221] showed that just 56% of individuals had a lipid measurement taken before starting treatment. One would anticipate that future audits are likely to show that this percentage has increased as the importance of monitoring has become recognised. Previously, for example, total cholesterol levels would be more likely to be measured if a patient was thought to have a high risk of dyslipidaemia and, as a consequence, rates may appear to be artificially high. Thus, as the percentage tested increases any potential selection bias will decrease and over time levels may be observed to decrease as a greater proportion of patients who are not at a particularly high risk are also routinely monitored.

Another consequence of the increased awareness of HAART-related toxicities is that the frequency of monitoring of laboratory markers has increased in those on treatment and the time interval between consecutive measurements has shortened. As a result, adverse events are likely to be diagnosed sooner and at an earlier stage allowing for interventions to be made earlier.

As new drugs are introduced with better toxicity profiles, it may be that the drugs or drug combinations used to treat HIV in earlier periods are less widely used to treat new patients. Thus, the relevance of findings from early studies may be questioned and the incidence of disorders that we could expect to see in the future may be overestimated. However, some individuals may still be receiving the earlier licensed HIV drugs, although probably as second line and later HAART therapy or they may have been receiving them for a while and therefore may have self-selected as being less likely to have a toxicity. Whilst clearly this represents good clinical practice, this may lead to an apparent increase in rates of side effects as a consequence of an increase in the frequency of monitoring, rather than a real increase in the number of events.

Conversely, patients with measurements in earlier years may be those at the greatest risk of developing a toxicity. Thus, once testing is routinely introduced there may appear to be a decrease as patients included are no longer those at the greatest risk of experiencing a toxicity.

There are several other factors that have changed over time that could lead to the appearance that the incidence rates of HAART-related toxicities have decreased. For example, greater awareness of toxicities for patients and clinicians and the introduction of better supportive therapy may all lead to decreases in the rates of toxicities observed. Therefore, it is particularly important to account for and be aware of potential biases associated with calendar time when studying HAART-related toxicities. These issues are investigated in Chapter 4.

2.6 Methodological Issues: Differences in demographic and treatment factors between studies

As mentioned previously, the demographic profiles of different patient populations may limit the comparison of the rates of disorders between studies. Whilst demographic characteristics of those patients included in studies are nearly always presented, the effect that these differences would have, if any, on the incidence rates reported is not always immediately obvious. For example, it is not clear whether women would be more likely to develop hepatotoxic disorders than men, and so populations containing very different proportions of men and women may not be comparable. However, consistent results obtained from a series of studies carried out in very different demographic populations would support evidence of a real effect due to antiretroviral treatment rather than an indirect effect due to some unknown confounding factor [222]. Thus, heterogeneity in study populations may act as both a help and a hindrance. One potential source of heterogeneity, the CD4 cell nadir, is investigated in Chapter 6.

Another factor related to demographic information is the treatment history of those in studies. An RCT that studies the benefits of switching from regimen A to regimen B compared to remaining on A could lead to entirely different results to a RCT of previously ART-naïve patients randomised to regimen A or regimen B. The former is examining the effect of the regimen B *after* the first regimen had been taken compared to remaining on regimen A, whereas the latter situation compares regimen A and regimen B when they are both taken as first-line regimens. Any residual effect of regimen A on those who switch treatment regimens will be evident in the first situation, which will not happen in the second situation. Furthermore, many adverse events occur

in the first few months on a regimen. Thus, in the first trial lower rates of adverse events for drug A may be seen than in the second situation. The first situation would also not include those individuals who started A, developed an adverse event quickly and then stopped the regimen. There are several different ways to adjust for previous treatment regimens when carrying out an analysis but this may not solve all biases, particularly when patient populations are quite different. For example, analyses can adjust for whether an individual has ever been exposed to a specific antiretroviral or a class of antiretrovirals, the duration of exposure to a specific antiretroviral or class, the number of drugs ever exposed to, or the recent or current use or a specific antiretroviral of drug class. Each may impact slightly differently on the results obtained, particularly when antiretroviral histories are complex, and the interpretation of each will be different.

The exact antiretroviral regimens of individuals included in studies may also lead to differences in the incidence and prevalence rates quoted for metabolic disorders. For example, Costa et al [216] described changes in total cholesterol and triglycerides amongst those on rescue therapy HAART regimens (those that included at least 5 antiretrovirals including amprenavir and ritonavir). Rakatoambinina et al [215] described changes in total cholesterol levels amongst those on protease inhibitor-containing HAART regimens including both previously-naïve and experienced patients. Besides the fact that the previous treatment history of the individuals included in the two studies was very different, if different PIs have different toxicity profiles [198;217] then the increases in cholesterol amongst those in Costa's study would not necessarily be expected to be comparable to those observed in Rakatoambinina's study both because of the number and type of PIs included in each regimen. Thus, although both studies compare 'PI-containing' regimens, the incidence of toxicities in each would be expected to be quite different. This extends to other studies, where some authors have studied everyone on combination therapy including non-nucleoside reverse transcriptase inhibitors [213;223], or just those receiving a protease inhibitor (PI) [224] or other designs and combinations [225].

2.7 Methodological Issues: Conclusions

There are several issues that must be taken into account when studying the side effects of HAART between different studies, whether RCTs or observational studies. Differences in definitions of the study outcomes, whether the comparison is being

made in a RCT or an observational study, the HAART regimens being administered, the calendar date of the study, and the demographic factors of the population being studied can all affect the incidence rates quoted, and so may be able to explain some of the differences in incidence rates observed between studies. The following sections look at the issues that are specific to different disorders that may explain the differences in incidence and prevalence of HAART-related toxicities observed.

2.8 Metabolic disorders

2.8.1 Introduction

Several papers have examined the effects of HAART on metabolic disorders [180;201;205-213;213-219;223-230]. There are also several reviews examining this issue [170;231-239], and recommendations for appropriate treatment for these disorders [240]. In particular, two disorders have been extensively studied. The first is lipodystrophy, characterised by fat redistribution in the body. This can take the form of either *lipoatrophy*, known as fat wasting, or *lipohypertrophy*, which is fat accumulation. The second disorder is dyslipidaemia (demonstrated by abnormal total cholesterol, triglycerides, low density lipoprotein (LDL)- and high density lipoprotein (HDL)-cholesterol levels) as a surrogate for cardiovascular clinical events (such as myocardial infarctions (MI), strokes and death). There is some evidence linking lipodystrophy to both cardiovascular disease [241-243] and dyslipidaemia [244] and many studies have examined both disorders [205;207;213;215;218;219;224]. Therefore I have included both disorders in this section. Table 2.1 summarises the most recent studies that have investigated antiretroviral-related metabolic disorders with at least 100 patients included. The following sections highlight a few points that could possibly explain why differences are observed in these studies with regard to the rate of metabolic disorders.

2.8.2 Cardiovascular endpoints

There is some evidence that HAART is protective against the occurrence of cardiovascular disease and that immune competence is associated with increases in lipid levels [245;246]. Nonetheless, a number of studies have investigated the association between cardiovascular disease and HAART. The most relevant outcome measures when evaluating the prevalence and incidence rates of HAART-related metabolic disorders are the occurrence of clinical events, such as strokes, MIs and death [195;201;228;247;248]. As mentioned previously, these events tend to have low incidence rates and so large numbers and person-years of follow up are required (all of these studies

include at least 4000 individuals). The findings from these studies are conflicting. Studies by the D:A:D study group [195], and Mary-Krause et al [247] found an increase in the incidence of cardiovascular disease as the amount of time of exposure to HAART increased. In contrast, the studies by Bozzette et al [228] and Klein et al [201] found conflicting results. However, the latter two studies both investigated the incidence of MIs or coronary heart disease in the HAART time period compared to the non-HAART time period, rather than by the length of each individual's exposure to antiretroviral therapy. Braitstein et al [248] used a similar outcome measure to Bozzette et al, examining the rate of interventional cardiac procedures by calendar year, and, in contrast to Bozzette et al, found a significant increase in the rate over time. This could be confounded by an increased willingness to perform cardiovascular procedures in the HAART era. However, the authors only studied those actually receiving antiretroviral therapy, rather than all HIV-positive individuals as was done in Bozzette's analysis. The final study, by Coplan [227] et al, found that the incidence of MIs was statistically higher amongst those receiving PI-containing HAART compared to those receiving NRTIs alone. Clearly, different investigators have also used different outcome measures (such as cardiovascular disease, cerebrovascular or cardiovascular disease, interventional cardiac procedures) and this could also potentially explain some of the differences found between these studies.

The studies assessing the prevalence and incidence of dyslipidaemia also give conflicting results [249]. In general, they need to be examined in two groups: studies that look at the proportion who have a total cholesterol or triglyceride measurement above a certain threshold [205;207;209;213;218;224;250;251], and studies investigating changes in these markers [180-182;206;217;229;249;252]. Amongst those studies looking at the prevalence of total cholesterol levels of more than 6.2 mmol/l there is great variation in the rates observed. The most obvious reason is the different treatment regimens that were used in the different studies and the threshold used to define hypercholesterolaemia. Amongst a general HIV positive population [253] (both treated and untreated) a prevalence rate of 21% was observed. Collazos et al [209] found similar incidence rates of raised total cholesterol levels amongst HIV positive, but ART-naïve, individuals of 19% compared to an incidence of 33% amongst those receiving ART. The rates of hypercholesterolaemia in an RCT after 12 months of therapy reported by Martinez et al [251] were 7% in the abacavir arm, 21% in the nevirapine arm, and 29% in the efavirenz arm. Tsiodras et al [218] found a prevalence of hypercholesterolaemia after 5 years of HAART in previously ART-naïve individuals of 24%. Clearly, the current treatment regimens and the treatment histories of these studies are very different. Two studies used a threshold of 5.5 mmol/l to define hypercholesterolaemia, and found very

different results. Carr et al [207] found the incidence rate was 11% amongst PI-naïve individuals compared to 58% amongst PI-experienced individuals. Heath et al found 16% of those receiving ART had hypercholesterolaemia. Another potential source for differences in observed prevalence rates between these studies is the proportion of men included in studies. Although most studies include predominantly men, the percentage varies from 39% in the study by Martinez et al to 99% and 100% in the studies by Carr et al and Collazos et al respectively. In contrast, there is less variation in the average age of individuals included in these studies, with all studies having a mean or median age of between 35 to 40 years.

Studies investigating absolute changes in total cholesterol also report contrasting results. Of those studies comparing changes relative to a switch in an antiretroviral regimen, Matheron et al [182] found in a prospective cohort study that 48 weeks after a switch had taken place those receiving combivir (AZT/3TC) and abacavir had an increase of 0.01 mmol/l compared to 0.53 mmol/l amongst those receiving combivir and nelfinavir. This is in agreement with the study mentioned previously by Martinez et al [251], which found a lower prevalence of hyperlipidaemia amongst those receiving abacavir and other NRTIs without a PI compared to those receiving PI-containing HAART. Consistently, Ruiz et al [252] found that 48 weeks after individuals were randomised to change a PI to Nevirapine, the mean total cholesterol fell from 228 mg/dl (5.9 mmol/l) to 207 mg/dl (5.4 mmol/l), whereas the mean value amongst those who remained on their PI-containing regimens changed only slightly from 222 mg/dl (5.74 mmol/l) to 220 mg/dl (5.69 mmol/l). All individuals in this study were experiencing lipodystrophy at baseline, which may mean that they had particularly elevated total cholesterol levels, and may go some way to explaining the greater decrease in total cholesterol in the non-PI containing arm compared to that reported in the non-PI containing arm in the study by Matheron et al.

Joly et al [180] also studied changes in total cholesterol amongst those who were treatment experienced, but who had only been previously exposed to NRTIs. Changes in total cholesterol were examined at a much later time point (30 months), but no significant changes were found in either arm (d4T/3TC/IDV or AZT/d4T/IDV), perhaps either reflecting that changes in TC were not sustained in the long-term, or that the addition of indinavir did not further raise TC levels above those observed amongst those receiving NRTIs. Periard et al [217] compared changes in TC to pre-therapy levels after a mean of 470 days of HAART. The authors found that those with PI-containing therapy had greater increases (RTV arm 2.0, IDV 0.8, NFV 0.2 mmol/l) compared to those who were PI-naïve (0.11 mmol/l). However, the changes they found were much

larger in the RTV- and IDV- containing arms than those reported by Ruiz and Matheron. This may be because all individuals were previously antiretroviral-naive, and so changes were much greater. Van Leth et al [181] studied previously-naïve individuals starting HAART in an RCT, and found results contrasting to those of Periard's cohort study. After 48 weeks of HAART, the mean increases in TC in mmol/l were 1.01 in the once-daily nevirapine, 0.95 in the twice-daily nevirapine arm, 0.37 in the efavirenz arm, and 0.45 in the efavirenz plus once-daily nevirapine arm. All arms were PI-naïve, yet increases were much greater than those seen in the corresponding arm in Periard's study. Visnegarwala [249] studied a population who were similarly previously-naïve, and found similar increases in total cholesterol levels in the PI-containing and PI-naïve arms (15.5 mg/dl [0.4 mmol/l] in the PI-arm, compared to 18.0 mg/dl [0.5 mmol/l] in the PI-sparing arm), contrasting with Periard's study. The above studies have similar characteristics in terms of median age and proportion of men.

Different demographic profiles of patients included in different studies could further explain some of the differences observed in the rates of cardiovascular endpoints. It has been shown that, in general, HIV positive populations possess many traditional risk factors associated with increased chances of cardiovascular disease in HIV negative populations, such as increased body mass index, increased rates of smoking and alcohol consumption and high blood pressure [205;254-257]. These data are often not available in many cohorts or RCTs, and so cannot be adjusted for.

Similar contradictions exist when examining changes in triglycerides and in the incidence of hypertriglycerideaemia. A further problem that can lead to discrepancies with this endpoint is that triglycerides are affected by whether the patient was fasting or not at the time of measurement. There is some evidence that non-fasting triglycerides are of more interest [258], but it is important that studies indicate which method was used as this will have an effect on both the range of values quoted, plus the percentage identified as having dyslipidaemia. Most studies report using fasting triglycerides measures [181;182;217;218;224;249;251;252], some non-fasting [259], some both [205], and others have not stated whether triglycerides were taken under fasting conditions or not [180;206;209;213;229;250;253].

2.8.3 Lipodystrophy

There are also discrepancies amongst studies reporting incidence rates of lipodystrophy as no clear case definition exists as yet [260]. Methods that have been used to define lipodystrophy include patient assessment [207;214;218], clinician assessment [207;218]

and the use of dual-energy x-ray absorptiometry (DEXA) scans [210]. It has been shown that there is great disparity between these measures [260;261]. Furthermore, Yang et al demonstrated that different DEXA scans from different manufacturers can provide disparate results [262]. As patient and clinician defined lipodystrophy are subjective measures, the rates observed are likely to be strongly affected by how important the assessor considers lipodystrophy to be. Recently, Carr et al [263] and Ioannidis et al [264] have attempted to establish a clear case definition and so we must wait for studies to be carried out using this new definition in order that results from different studies can be compared. Also, some studies define lipodystrophy as one outcome, whereas others investigate lipoatrophy, lipohypertrophy and mixed lipodystrophy separately, and this will also affect the rates observed, particularly as HAART may have differential effects on each. Furthermore, some studies consider individuals to only have either lipoatrophy or lipohypertrophy [213] (they sometimes also include a mixed lipodystrophy category [250]), whereas other studies allow individuals to have both lipoatrophy and lipohypertrophy [180].

Three studies have examined patient-defined lipoatrophy (LA) and lipohypertrophy (LH). Martinez et al [251] found in an RCT that the rates of lipoatrophy were 34%, 32%, and 33% for arms containing abacavir, nevirapine and efavirenz respectively 12 months after switching from a PI. The rates of lipohypertrophy reported were 14%, 16% and 19% in the same arms. Heath et a l[213] found in a prospective cohort study that after 1 year of ART the incidence rates of lipoatrophy and lipohypertrophy were 27% and 21% respectively. In a different study by Heath et al, in which the rates of LA and LH were investigated after 12 months of HAART (as opposed to any ART regimens) in those who were previously antiretroviral naïve, rates of 29% and 23% respectively were observed. Although Martinez's RCT had slightly higher rates of lipoatrophy and slightly lower rates of lipohypertrophy, overall the rates were comparable between the latter two observational studies. One study (Carr et al [207]) looked at patient-defined lipodystrophy (as one overall condition) after a mean of 20 months of HAART, and found 83% had LD in the PI-containing arm and 4% in the non-PI arm. This study found much lower rates of lipodystrophy in the PI-naïve arm than Ruiz's RCT where none of the arms included a PI. However, all arms in Ruiz's trial had been previously exposed to PIs, and this may go some way to explaining the differences observed.

Clinician-defined lipodystrophy was considered in two studies. Lichenstein et al [230] found in a prospective cohort study, that 62% of those receiving HAART had lipodystrophy, and a further 13% developed it in the next 20 months, which is much higher than the rates seen above. There may be other factors that could explain these

higher rates. For example, 59% of those in Lichenstein's study were older than 40 years, which is a slightly older population than that seen in the other studies and this could possibly have an impact on the rates observed. Van der Valk et al [183] found in an RCT that 8% in the RTV/SQV arm had lipodystrophy after 96 weeks compared to 24% in the RTV/SQV/d4T arm. Both arms had much lower prevalence rates than those seen in the patient-defined studies reported above, especially as both arms contained PIs.

Studies of patient- or clinician-defined lipodystrophy also reported a lot of variation in the observed rates of lipodystrophy. Rakatoambinina et al [215] found that 29% of those on PI-containing HAART developed LD after a mean of 20 months on HAART. Joly et al [180], however, found in an NRTI-experienced population after 30 months of HAART rates of LA of 70% in those receiving d4T and 43% in those receiving AZT (both arms received a PI also). The rates of LH observed were 51% and 37% respectively. These rates are clearly much higher than those observed by Rakatoambinina. In contrast, Tsiodras et al [218] found that the cumulative incidence of lipodystrophy after 1 year of HAART in 221 previously naïve individuals was just 13%.

Table 2.1 – Summary of studies examining the relationship between HAART and metabolic disorders

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
Studies assessing incidence of cardiovascular events								
Klein[201], USA	Retrospective cohort	4159	n/a	42	ART-period (up to 1996) and non-ART period	-	Hospitalisation for CHD events	CHD: 72, No difference between ART- and non-ART period
Mary-Krause[247], France	Prospective cohort	19795	100	n/a	PI-containing regimens	-	Incidence of MI with <18 months (group 1), 18-29 months (group 2) or >30 months (group 3) of PI exposure	Group 1: 8.9 per 10000 pyrs Group 2: 19.2 per 10000 pyrs Group 3: 34.7 per 10000 pyrs
Bozzette[228], USA	Retrospective cohort	36766	98	[a]	70% on ART	-	Incidence rates of CVD or CBD	Between 1995 and 2001 decreased from 1.7 to 0.9 per 100 pyrs.
Coplan[227], USA	RCT meta-analysis	10986	86	37	PI-containing or NRTI-only	-	Incidence of MIs	PI-containing: 1.82 /1000 pyrs NRTI-only: 1.05 /1000 pyrs
Braitstein[248], Canada	Retrospective cohort	5082	n/a	n/a	All ART-experienced	-	Incidence of interventional cardiac procedures by calendar year	1995; 5.25; 1996 1.23; 1997 0.39; 1998 1.57; 1999 3.54; 2000 6.18 per 1000 pyrs
Iloeje [265], USA	Prospective cohort	7542	86	39	PI-treated or PI-unexposed, followed for median of 3 years	-	CVD disease	CVD: PI-exposed=9.8 per 1000 pyrs; PI-unexposed=6.5 per 1000 pyrs
D:A:D study[195], worldwide	Prospective cohort	23468	76	39	80% ART experienced	-	Incidence of MIs	23% increase per additional year of exposure to HAART
Phillips , worldwide[196]	RCT	5472	73	44	Randomised to continuous ART (VS) or ART	-	Incidence of major CVD event	57% increase in hazard for DC arm vs VS 12% increase per additional year of

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
					interruption strategy (DC)			PI exposure in DC arm
D:A:D study[266], worldwide	Prospective cohort	23437	76	39	By end of follow-up 80% exposed to PIs and 64% exposed to NNRTIs	-	Incidence of MIs	PI exposure: 11% increase per additional year of exposure NNRTI exposure: 2% increase per additional year of exposure
Studies investigating dyslipidaemia								
Friis-Moller,[205] worldwide	Cross-sectional study	17852	76	39	13% ART naïve	-	TC>6.2 mmol/l	TC: 21%
Visnegarwala [249], USA	Prospective cohort	79	65	38	Previously naïve, 52% PI-containing and 48% PI-sparing regimens	-	Increases in TC, TG after 6 months compared to pre-ART	TC: PI-arm=15.5; PI-sparing=18.0 mg/dl TG: PI-arm=90; PI-sparing=0.9 mg/dl
Koppel[229]	Observational cohort	409	n/a	42	PI-containing HAART and ART-naïve	-		TC: PI arm=6.0; Naïve arm 4.6 mmol TG: PI arm=2.7; Naïve arm 1.5 mmol
Collazos[209], Spain	Cross sectional study	197	100	37	15% ART naive	-	TC >200 mg/dl or TG>200 mg/dl	33% receiving ART vs. 19% ART naïve
Thiebaut[253], France	Prospective cohort	1429	74	36	PI-containing, NNRTI-containing, no ART	-	Factors associated with increases in TG	PI use: 21% increase
Matthews[259], UK	Cross sectional study	135	91	38	First line HAART	-	Factors associated with raised TC, TG	TC: age and TG associated TG: TC associated
Van Leth[181], worldwide	RCT	1216	n/a	n/a	Previously naïve, randomised to NVP (od), NVP (bd),	-	Mean Changes in TG, TG from baseline to 48 weeks	TC: NVP (od) 1.01; NVP (bd) 0.95; EFV 1.11; NVP+EFV 1.41 mmol/l TG: NVP (od) 0.20; NVP (bd) 0.05;

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
					EFV, NVP (od)+EFV with 3TC+d4T			EFV 0.37; NVP+EFV 0.45 mmol/l
Periard[217], Switzerland	Prospective cohort	121	73	40	PI-containing and PI-naïve	-	Mean changes from pre-treatment levels in TC and TG after an average of 470 days	TC: RTV 2.0; IDV 0.8; NFV 0.2; PI-naive 0.1mmol/l TG: RTV 1.83; IDV 0.26; NFV 0.26; PI-naïve 0.11mmol/l
Domingo [267], Spain	RCT (results from single arm)	94	70	n/a	Heavily pre-treated, switching d4T for TFV	-	Mean TC and TG levels at baseline and 12 weeks	TG: baseline 458; 12 weeks 279 mg/dl TC: baseline 266; 12 weeks 231 mg/dl
Gathe [268], Worldwide	RCT	649	73	36	Naïve, starting FAP/ABA/3TC or NFV/ABA/3TC	-	Median TG, TC, LDL, HDL increase at 48 weeks	TG FAP: +58; NFV=+41 mg/dl TC, LDL HDL; similar increases in both groups
Llibre [269], Spain	Prospective cohort	352	71	41	Patients switched from d4T to TFV and with 48 weeks of lipid follow-up	-	Median TC, TG, LDL and HDL at baseline and 48 weeks	Baseline: TG=180; TC=206; HDL=42; LDL=122 mg/dl 48 weeks: TG=145; TC=188; HDL=41; LDL=114 mg/dl
Calza[270], Italy	RCT	142	60	39	PI-containing regimen with TC>250 mg/dl and TG>200 mg/dl. (1) Switch to NVP; (2) Switch to EFV; (3) Receive pravastatin; (4) Receive bezafibrate	-	Mean change in TG, LDL, HDL and TC at 12 months in mg/dl	TC: (1) -27.1; (2) -10.2; (3) -45.8; (4) -37.6 TG: (1) -25.2; (2) -9.4; (3) -41.2; (4) -46.6 LDL: (1) -25.2; (2) -8.7; (3) -39.6; (4) -35.1 HDL: (1) +3.1; (2) +1.9; (3) +10.2; (4) +7.7
Young [271], Switzerland	Prospective cohort	1065	68	37	Starting HAART with at least one subsequent lipid measurement	-	Mean change in TC, HDL, TG per year of use	TC: PI=+0.18; NNRTI=+0.21; d4T=-0.10; ABA=-0.11 mmol/l HDL: PI=+0.01; NNRTI=+0.10; d4T= +0.00; ABA=+0.02 mmol/l

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
								TG: PI=+0.21; NNRTI=-0.13; d4T=+0.04; ABA=-0.20 mmol/l
Jones[272], UK	Prospective cohort	1664	86	n/a	First line HAART with baseline TC<5.5 mmol/l	-	TC>6.5 mmol/l	10%
Van Leth [273], Worldwide	RCT	706	n/a	n/a	Naïve and randomised to EFV or NVP with d4T and 3TC	-	Mean percentage change in TC, TG, LDL and HDL after a median of 84 weeks	TC: NVP=+45%; EFV= +45% TG: NVP=+95%; EFV=+126% LDL: NVP=+65%; EFV=+72% HDL: NVP=+34%; EFV=+31%
Richter [274], USA	Retrospective cohort	900	84	40	All patients at clinic with mean follow up 3.3 years	-	Any of: TC>240 mg/dl; LDL >160 mg/dl; TG>200 mg/dl	53.8%; more common amongst PI treated
Shikuma [275], USA	RCT	1147	n/a	n/a	Naïve patients randomized to (1) ZDV/3TC/EFV (2) ZDV/3TC/ABA/EFV or (3) ZDV/3TC/ABA	-	Mean changes at week 24 in TC, LDL, HDL and TG	TG: (1)=7; (2)=18; (3)=-1 mg/dl TC (1)=23; (2)=28; (3)=5 mg/dl LDL: (1)=9; (2)=14; (3)=1 mg/dl HDL: (1)=10; (2)=10; (3)=5 mg/dl
Fontas [276], Worldwide	Cross sectional study	7483	76	38	Antiretroviral naïve or receiving first cART regimen	-	Median TC, HDL, LDL, TG levels	TC: naïve=4.4; single PI=5.3; dual PI=5.7; NNRTI=5.1 mmol/l HDL: naïve=1.1; single PI=1.1; dual PI=1.1; NNRTI=1.3 mmol/l LDL: naïve=2.9; single PI=3.6; dual PI=3.8; NNRTI=3.2 mmol/l TG: naïve=1.3; single PI=1.8; dual PI=2.5; NNRTI=1.3 mmol/l
Hadigan[206], USA	Matched case-control	284	62	42	Cases: 71 patients with LD; All on ART. Controls: 213 HIV +ve patients without LD. 0% on	-	Mean TC and log TG levels	TC: 229 mg/dl in cases vs. 195 controls, log TG: 5.5 in cases vs. 4.7 controls

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
					ART			
Studies investigating lipodystrophy in those on HAART								
Lichenstein[230], USA	Prospective cohort	546	n/a	[b]	Already on HAART	Clinician reported	Prevalence of LD	LD: 62% at baseline, 13% further in next 20 months
Rakatoambinina[215], USA	Cohort	175	72	40	58% NRTI-experienced, receiving PI-containing HAART	Clinician + patient reported	Prevalence of LD	29% developed LD after a mean 20 months on HAART.
Van der Valk[183], Netherlands	RCT	175	n/a	n/a	PI, d4T naïve; RTV/SQV with or without d4T	Clinician reported	Prevalence of LD at 96 weeks	RTV/SQV 8%; RTV/SQV/d4T 24%
Carr[225], Australia	Cross sectional study	195	99	40	26% on PI and 16% PI naïve	Clinician + patient reported	Prevalence after a mean of 27 months	64% of PI and 3% PI naïve
Young [277], Switzerland	Prospective cohort	675	69	38	Naïve starting HAART	Patient reported	Incidence of LD	LD : 13.2 per 100 person years
Seminari[278], Italy	Cross sectional study	504	72	38	All on ART undertaking screening between January-June 2000	Marrakech score-clinician reported	Prevalence of LD	40%; compared to no LD were more likely to have more ART, PI and ZDV use
Karmon [279], USA	Cross sectional study	161	0	40	Visiting clinic July-August 2000	Patient reported	Prevalence of LD	7.5%
Bernasconi [280], Switzerland	Cross sectional study	1359	73	[c]	Patients receiving ART in August-September 1999	Clinician + patient reported	Prevalence of LD and LA	LD: 43%; associated with longer d4T use LA: 28%; associated with longer d4T use LH: 30%; associated with longer d4T use

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
Cameron [281], USA	RCT	106	78	41	Randomised to LPV+3TC+ZDV followed by LPV monotherapy or to EFV+3TC+ZDV	DEXA	LA: >20% decrease in limb fat at 96 weeks LH: >20% increase in trunk fat at 96 weeks LD: LA and LH	LA: LPV 5%; EFV 34% LH: LPV 45%; EFV 44% LD: LPV 0%; EFV 16%
Tien [282], USA	Prospective cohort	815	0	[40]	HIV positive (n=605) and HIV negative (n=210)	Patient reported	30 month cumulative incidence of LA and LH	Peripheral LA: HIV+= 27%; HIV-=13% Central LA: HIV+= 23%; HIV-=13% Peripheral LH: HIV+= 28%; HIV-=31% Central LH: HIV+= 18%; HIV-=25%
Bonfanti [283], Italy	Prospective cohort	1480	72	37	PI-containing ART	Clnician + patient reported	Incidence of LD, LA and LH per 100 person-months	LD: IDV 11.1; RTV 13.6; SQV 7.5; NFV 8.2; SQV/RTV 9.0 LA: IDV 5.1; RTV 7.4; SQV 4.8; NFV 3.4; SQV/RTV 5.5 LH: IDV 6.1; RTV 6.1; SQV 2.7; NFV 4.9; SQV/RTV 3.5
Studies investigating both lipodystrophy and metabolic abnormalities								
Martinez[251], Spain	RCT	460	39	75	Switch PI for NVP (n=155), EFV (n=156) or ABA (n=149)	Patient reported	TG>4.5 mmol/l; TC>6.2 mmol/l; incidence of LH and LA at 12 months	TG: ABA 4%; NVP 12%; EFV 13% TC: ABA 7%; NVP 21%; EFV 29% LA: ABA 34%; NVP 32%; EFV 33% LH: ABA 14%; NVP 16%; EFV 19%
Joly[180], France	RCT	170	81	37	Pre-treated with ddI, AZT or ddC receiving d4T/3TC /IDV or AZT/3TC/IDV	Clinician + patient reported	Changes in TC, TG and prevalence of LA and LH after 30 months	TC, TG: No significant differences LA: 70% d4T arm vs. 43% AZT arm LH: 51% d4T arm vs. 37% AZT arm

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
Heath, Canada[213]	Prospective cohort	745	88	40	Receiving ART, 87% on triple or more therapy.	Patient reported	Incidence of LA, LH, TC>5.2 mmol/l TG> 2.3 mmol/l after 1 year of ART	LA:27%, LH 21%, TC 16%, TG 10% Raised TC, TG associated with PI used and RTV LA, LH associated with d4T use
Carr[207], Australia	Prospective cohort	141	99	n/a	PI-containing and PI-naive arms.	Patient reported	LD; TC >5.5 mmol/l: TG >2.0 mmol/l: HDL<0.09 mmol/l after a mean of 20 months	LD: 92 (83%) PI vs. 1 (4%) non-PI TG: 50% in PI vs. 22% non-PI TC: 58% PI arm, vs. 11% non-PI arm
Tsiodras[218]	Retrospective cohort	221	77	37	All previously-naïve, 45 remained PI-naive	Clinician + patient reported	Prevalence of TC>6.2 mmol/l, TG>5.6 mmol/l and LD at 5 years	TC: 24%, TG 19%; LD 13%
Paparizos[224], Greece	Retrospective cohort	324	90	35	2 NRTI+PI	Unknown	Incidence of LD TG>150 mg/dl	LD: 11.6% after a mean of 4.7 months TG: 52% after a mean 10 months.
Ruiz[252], Spain	RCT	106	75	39	All with LD: remain with PI-containing or switch to NVP	DEXA scan	Changes in LD, TG, TC at 48 weeks	LD: No significant changes TC: 228 to 207 mg/dl in NVP arm, 222 to 220 mg/dl in PI arm TG: 270 to 217 mg/dl in NVP arm, 285 to 270 mg/dl in PI arm
Matheron[182]	RCT	195	n/a	n/a	Switch to COM/ABA or COM/NFV	Unknown	Median increase in TC and TG at 48 weeks; prevalence of LD at 48 weeks	TC: COM/ABA 0.01 mmol/l; COM/NFV 0.53 mmol/l TG: COM/ABA 0.01 mmol/l; COM/NFV 0.17 mmol/l LD: COM/ABA 6%; COM/NFV 12%
Heath[250], Canada	Prospective cohort	366	89	38	Previously-naïve; all HAART regimens	Patient reported	Incidence of LA, LH, mixed LD, raised TG/TC after 12 months	LH:29%; LA 23%; Raised lipids: 9%: mixed LD 13%

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
Rodriguez-Delgardo, Spain[284]	Prospective cohort	90	63	38	Previously naive and >95% adherent	Clinician + patient reported	TC>240 mg/dl and TG>200 mg/dl	TC: 39% after a mean 248 days TG: 42% after a mean of 229 days LD: 40.7% after a mean of 488 days High TG levels associated with presence of LD at 6 months
Dube [285], USA	RCT	334	86	36	Naïve patients randomised to either ddI/d4T or AZT/3TC and to one of EFV, NFV or EFV+NFV	DEXA scan to calculate limb fat	Median changes in TC at week 64 Median percentage change in total limb fat	TC: NFV≈+31; EFV≈+31 ZDV/3TC≈+30; ddI/d4T≈+50 LD: NFV=-13.1; EFV=+1.8 ZDV/3TC=-16.8; ddI/d4T=+4.0
Moyle[286], UK	RCT	105	90	43	Switch from thymadine analogue to TFV or ABA	DEXA scan to measure limb fat	Mean change from baseline to 48 weeks in limb fat, TC, HDL, LDL, TG	LD: TFV=+329g; ABA=+483g TC: TFV=-0.45; ABA=+0.21 HDL: TFV=-0.11; ABA=+0.1 LDL: TFV=-0.25; ABA=+0.09 TG: TFV=-0.33; ABA=+0.07
Domingo [287], Spain	Cross sectional study	150	79	39	Receiving first line d4T/3TC/IDV or ZDV/3TC/IDV	Clinician + patient reported	Median TC, TG, LDL, HDL levels and prevalence of LH and LA	TC: d4T=5.0; ZDV=4.5 TG: d4T=1.6; ZDV=2.0 LDL: d4T=3.0; ZDV=2.6 HDL: d4T=1.0; ZDV=0.8 LH: d4T=26.7%; ZDV=17.3% LA: d4T=17.3%; ZDV=9.3%
Podzamczar [288], Spain	RCT	257	76	38	Naive patients randomised to d4T or ABA with 3TC and EFV	Clinician + patient reported	Presence of LA and LH at 96 weeks Mean change in TC, TG, LDL and HDL at 96 weeks	LA: ABA=4.8% d4T=38.3% LH: ABA=10.7% d4T=24.7% TC : ABA=1.09 ; d4T=0.89 mmol/l LDL : ABA=0.56 ; d4T=0.43 mmol/l HDL : ABA=0.47 ; d4T=0.20 mmol/l TG : ABA=0.28 ; d4T=0.85 mmol/l

First author and country	Study type	n	% male	Average age (yrs)	Treatment information	LD definition	Primary Endpoint	Main result
Haubrich [289], USA	RCT	693	80	38	Naive patients randomised to EFV+LPV/r or EFV+2NRTIs or LPV/r+2 NRTIs	DEXA scan	LA: >20% increase in extremity fat at 96 weeks Median change in TC, TG, HDL and non-HDL at 96 weeks	LA: EFV 32%; LPV 17%, both 9% TC: EFV +33; LPV +33; both +57 mg/dl TG: EFV +19; LPV +46; both +62 mg/dl HDL: EFV +9; LPV +8; both +16 mg/dl Non-HDL: EFV +22; LPV +26; both +44 mg/dl

TC=total cholesterol; TG=triglycerides; LDL=LDL cholesterol; HDL=HDL cholesterol; LD=lipodystrophy; LA=lipoatrophy; LH=lipohypertrophy CHD=coronary heart disease; CVD=cardiovascular disease; CBD=cerebrovascular disease; MI=myocardial infarction; VL=HIV RNA viral load; CD4=CD4 cell count; pyrs=person-years IQR=inter-quartile range; NS=not significant; ART=antiretroviral therapy; HAART=highly active antiretroviral therapy; med.=median; PI=protease inhibitor; NRTI=nucleoside reverse transcriptase inhibitor; NNRTI=non nucleoside reverse transcriptase inhibitor; ABA=abacavir; COM=combivir; ddI=didanosine; ZDV (or AZT)=zidovudine; ddC=zalcitabine; 3TC=lamivudine; d4T=stavudine; TFV=tenofovir;

ddI=didanosine; NVP=nevirapine; EFV=efavirenz; RTV=ritonavir; SQV=saquinavir; IDV=indinavir; FAP=fos-amprenavir; NFV=nelfinavir

[a]71% were aged between 35 and 55 years, but no mean or median age was given
[b]59% were aged over 40 years, but no mean or median age was given
[c]28% were aged <35 years, but no mean or median age was given
[d]52% were aged over 40 years, but no mean or median age was given

2.8.4 Conclusions

Clearly, very different results are obtained between studies when examining the incidence and prevalence of cardiovascular complications, dyslipidaemia and lipodystrophy. The most immediately obvious reasons for these differences are the treatment histories, current treatment regimens and length of time of exposure to HAART of those included in the studies. However, even amongst those studies where it may be reasonable to suppose that similar results might be expected very different results are obtained. This may be as a result of (in the case of lipodystrophy) using a subjective outcome measure, the type of study (RCT vs. observational study), or other unknown demographic factors associated with metabolic disorders, about which information is rarely available.

2.9 Hepatotoxic disorders

The mechanism of action by which antiretrovirals could cause hepatotoxicity is not yet known, and so it is still possible that any hepatotoxic effects seen after starting HAART could be caused by the immune reconstitution that occurs as a result of HAART rather than a true toxicity of the medications [290-292]. Furthermore, it has been noted that any potential deleterious impact of highly active antiretroviral therapy (HAART) on liver function, mainly among patients coinfected with HCV, should be balanced with the growing evidence supporting a beneficial effect of HAART on liver disease progression in this population [293]. Several studies [294-308] have investigated the relationship between hepatotoxicity and HAART and have observed differences in the strength of an association between the two. These studies are summarised in Table 2.2. There is great variety in the incidence rates of hepatotoxicity observed amongst those starting HAART. For example, Puoti et al [294] studied 755 individuals starting HAART and found an incidence rate of hepatotoxicity of 17 per 100 person years of follow-up. In contrast, Bonfanti et al [300] followed 1477 individuals starting PI-containing HAART and found an incidence rate of hepatotoxicity of only 2.7 per 100 person years of follow-up.

In a similar way to metabolic disorders, a major problem when assessing hepatotoxicity is the choice of definition used. Clinical endpoints are rarely used in this instance with the most common method used to define hepatotoxicity being the occurrence of an aspartate aminotransferase (AST) measurement and/or an alanine aminotransferase (ALT) measurement above a certain threshold. This threshold is usually a multiple of the upper limit of normal (ULN). The US-based AIDS Clinical Trials Group (ACTG) [309] define severe hepatotoxicity as either a grade 3 (5.1-10 x ULN) or grade 4 (>10 x ULN)

toxicity and these definitions, or an amended version, are often quoted in other studies [294;307].

Furthermore, the value taken as the ULN is not the same in all studies. The ULN for AST levels varies from 35 [310] to 47 [302] IU/L and from 31 [310] to 40 [302] IU/L for ALT levels. Although these differences are small, they could impact on the rates observed. For example Palmon et al and Wit et al carried out studies that were both retrospective, with similar characteristics in other respects, as shown in Table 2. There were some differences in the antiretroviral regimens, however all were receiving HAART. Nunez et al [299] used a cut-off of 35 IU/L for AST and 40 IU/L for ALT and observed a prevalence of 9% for hepatotoxicity, whereas den Brinker et al [302] observed a prevalence of 18% with ULNs of 47 IU/L and 37 IU/L for AST and ALT levels respectively. It is true that other, unknown factors could still lead to the differences observed between the two studies, but this may give an indication that the ULN chosen can impact on results observed.

It should be remembered that those with the highest levels at baseline are much more likely to experience hepatotoxicity under the above definition. Therefore, the definition of hepatotoxicity is often modified so that those who have ALT or AST levels above the ULN at baseline have to meet different criteria in order to be considered as having hepatotoxicity. This frequently involves requiring that these patients experience an ALT or AST level 2.5 times their baseline ALT/AST level. All of these can lead to apparent differences in the rates of hepatotoxicity observed.

Furthermore, some [299;301;307;310], but not all [295], studies exclude any hepatic endpoints that could have any occurred for non-drug related reasons. However, Wit et al and Den Brinker et al conducted studies with identical definitions of hepatotoxicity, except for the exclusion of cases with other causes. They also had similar proportions of those who were HCV positive and HBV positive, and used the same definitions for these two variables. The studies observed roughly similar hepatotoxicity rates: 18% in the Den Brinker study, and 14% in the Wit study. So, in this instance, this variable does not seem to have caused different results to be observed in these studies.

In addition to the factors mentioned in the methodological section, an important factor associated with hepatotoxicity is the presence or absence of either hepatitis B virus (HBV) or hepatitis C virus (HCV). The rates of HCV and HBV observed would be expected to vary greatly according to the means of transmission of HIV. Virtually all intravenous drug users and those infected with HIV through blood products would be

expected to be co-infected with HCV and in many cases (unless they had been vaccinated) with HBV, whereas those with a heterosexual or homosexual risk contact would have a much lower rate of infection [311]. As a result, the prevalence of HCV can vary dramatically from study to study, from 4% observed by Palmon et al [307] to 50% observed by Sulkowski et al [301], and this will have an impact on the rates of hepatotoxicity observed. The studies by Palmon and Sulkowski had similar definitions of hepatotoxicity (although Palmon excluded those with other causes, whereas Sulkowski did not) and the same definition for the ULN. Both studies included those receiving NNRTIs. However, Sulkowski et al reported prevalence rates of 16% for those on nevirapine and 8% for those receiving efavirenz, whereas Palmon et al observed a rate of only 1.1% overall.

The definition of HCV and HBV may also have an impact on observed rates of co-infection. Most studies define HBV on the basis of a positive surface antigen result, and sometimes [302;305], but not always [295;301], this must be measured on two consecutive occasions often six months apart. HCV is usually defined as a positive HCV antibody result [301], but sometimes the detection of HCV RNA is also considered a positive result [295]. Some patients are known to remain antibody positive despite clearing HCV [312], whereas others may have detectable HCV RNA but are not antibody positive [313]. Therefore, the definition of HCV and HBV positivity may affect the rates of HCV and HBV seen in a study.

As well as the potential limitations discussed in the methodological disorders section with regard to RCTs, there are particular additional limitations when assessing hepatotoxicity in clinical trials. Firstly, as shown by Madge et al [189], IDUs are often under represented in clinical trials and so too as a result are those with HCV or HBV. Furthermore, inclusion criteria often exclude those with elevated AST/ALT levels, and so the effect of HAART on developing hepatotoxicity amongst those who have the greatest chance of doing so is often not addressed in RCTs.

In a similar way to metabolic disorders there are a lot of factors that vary between studies and may go some way to explaining the differences in the rates of hepatotoxicity observed. The definition of hepatotoxicity, and the prevalence and definition of HBV and HCV can all impact on the results reported. Furthermore, differences between observational studies and RCTs, especially concerning the type of patients recruited to RCTs, can also have an effect on the rates observed.

Table 2.2 – Summary of studies investigating the relationship between HAART and hepatotoxicity

First author and country	Study type	N	IDU %	HCV definition and %	HBV definition and %	Demographic information	Definition of hepatotoxicity	ULN	Incidence/prevalence rate of hepatotoxicity
Puoti[294], Italy	Prospective cohort	755	65%	70% HCVAb	7% HBsAg	Previously ART-naïve	AST/ALT >5 times ULN, >2.5 times baseline if abnormal at baseline	Not stated	17 per 100 pyrs.
Saves[304], France	Prospective cohort	1047	17%	26% HCVAb	4% HBsAg	All starting PIs	ALT>5 times ULN. Excluded those with other causes	Not stated	5 per 100 pyrs (3.2, 7.6)
Gisolf[296], Belgium & Netherlands	RCT	208	5%	8% HCVAb	12% HBsAg	Previously naive starting RTV/SQV or RTV/SQV/d4T	ALT/AST>5 times ULN plus >100 U/L compared to baseline value	Not stated	18 (9%)
D'Arminio Monforte[297], Italy	Prospective cohort	1255	38%	47% HCVAb	7% HBsAg	Starting HAART, 214 ART-experienced	ALT>200 U/L or 5 times ULN or stopping regimen due to hepatotoxicity	ALT-40	61 (4.5%)
Aceti[298], Italy	Retrospective cohort	1325	50%	47% HCVAb	4% HBsAg twice >6 months apart	Receiving at least 1 PI for 6 months	ALT>5x ULN, or if high baseline then relative to baseline value	Not stated	44 (3.2%)
Nunez[299], Spain	Retrospective cohort	222	48%	38% HCVAb	5% HBsAg	Naïve starting HAART.	AST/ALT>5 times ULN or >3.5 times baseline if abnormal at baseline	AST-35, ALT-40	21 (9%)
Bonfanti[300], Italy	Prospective cohort	1477	48%	47% HCVAb	7% HBsAg	Starting PI containing ART	AST/ALT>5 times ULN, changes from baseline if abnormal baseline	Not stated	2.7 per 100 pyrs (2.6, 2.8)
Sulkowski[301], USA	Prospective cohort	298	54%	52% HCVAb	3% HBsAg	Starting ART	AST/ALT>5 times ULN or >3.5 times baseline if abnormal at baseline	AST-35, ALT-31	89 (29%)
Den Brinker[302], Netherlands	Retrospective cohort	394	9%	14% HCVAb	7% HBsAg twice >6 months apart	PI-naïve starting PIs	AST/ALT>5 times ULN, at least >100 from baseline	AST-47, ALT-37	70 (18%)

First author and country	Study type	N	IDU %	HCV definition and %	HBV definition and %	Demographic information	Definition of hepatotoxicity	ULN	Incidence/prevalence rate of hepatotoxicity
Martinez[303], Spain	Prospective cohort	610	35%	46% HCVAb	9% Not stated	Starting NVP-containing regimen (13% ART naïve)	AST/ALT >3 times increase from baseline	-	76 (12.5%)
Saves[295], France	Prospective cohort	(i)1253; (ii) 748	(i) 27% (ii)21 %	(i) 27% (ii) 35% anti-HCV or HCV-RNA	(i) 9% (ii) 13% HBsAg	(i) 2NRTI (ii)PI-containing HAART	ALT>200 U/L or 5 times the ULN.	Not stated	(i) 5.7 per 100 pyrs (ii) 7.3 per 100 pyrs
Wit[305], Netherlands	Retrospective cohort	560	7%	9% HCVAb	11% HBsAg twice >6 months apart	Previously naïve starting HAART	>5 times the ULN, >100 U/L compared to baseline value. Excluded those with other causes	AST-47, ALT-37	76 (14%)
Sulkowski[306], USA	Prospective cohort	568	42%	43% HCVAb	8% HBsAg on 2 consecutive occasions	Starting NNRTIs	AST/ALT>5 times ULN, >3.5 times baseline if abnormal at baseline	AST-35, ALT-31	40/256 (16%) on NVP 25/312 (8%) on EFV
Palmon[307], USA	Retrospective cohort	272	4%	12% HCVAb	9% HBsAg	Receiving NNRTIs	>5 times ULN, >3.5 times baseline if abnormal at baseline. Excluded those with other causes	AST-35, ALT-31	3/272 (1.1%)
Martin-Carbonero[314]	Retrospective cohort	42	88%	100% not stated	0% HBsAg	Previously naive starting HAART	>5 times ULN, >3.5 times baseline if abnormal at baseline	Not stated	6/42 (14%)
Macias, Spain [315]	Cross sectional study	152	86%	100% HCVAb	Not stated	41% naive, 59% received HAART (of whom 87% had received a PI)	Fibrosis stage ≥F3*	-	PI-containing: 49% NVP-based: 36% EFV-based: 42%
Sulkowski[316], USA	Prospective cohort	1161	43%	45% HCVAb	10% HbsAg	PI-naïve, Starting PIs	AST/ALT>5 times ULN, >3.5 times baseline if abnormal at baseline.	AST -35 ALT – 31	9.94 per 100 person years
Bonfanti[317],	Prospective	551	40%	40% HCVAb	7% HbsAg	Either naive or	Grade 3 event as	N/a	Naive: 0.54 / 100

First author and country	Study type	N	IDU %	HCV definition and %	HBV definition and %	Demographic information	Definition of hepatotoxicity	ULN	Incidence/prevalence rate of hepatotoxicity
Italy	cohort					experienced starting LPV/r	defined by ACTG		pyrs Experienced: 0.48 /100 pyrs
Mocroft [318], Europe	Prospective cohort	10937	24%	17% HCVAb	5% HbsAg	(1) All under follow-up in study (2) Those who had received HAART	Death from liver related disease	-	(1) Overall: 3.5 per 1000 pyrs (2) 12% increase in risk per year longer
Manfredi[319], Italy	Retrospective cohort	742	n/a	14% not stated	3% not stated	Receiving NVP or EFV ≥12 months and >90% adherence	>2 fold increase in AST/ALT levels	-	NVP: 92/346 (27%) EFV: 54/396 (14%) p<0.001
Meraviglia [320], Italy	Prospective cohort	782	42	39% HCVAb	47% persistent HbsAb	Treatment experienced starting LPV mean 373 days follow-up	>100 IU/L higher than baseline if abnormal at baseline	AST-40 ALT-40	9.1%
Maida [321], Spain	Cross sectional study	3200	n/a	0% not stated	0% not stated	All HIV positive patients in 2004	Persistently elevated LFTs in absence of other causes	AST-40 ALT-55	0.5%; risk factor is use of ddI
Sanne [322], South Africa	RCT	468	n/a	n/a	NVP: 4% HbsAg	Naïve receiving NVP or EFV and randomized to d4T/3TC or d4T/FTC	>5 times ULN	n/a	NVP: 66 (17%) EFV: 0 (0%)
Barreiro [323], Spain	Cross sectional study	1028	16	0 % serum HCV RNA	0% HBsAg/HBV DNA	All at clinic in 2005 without HBV, HCV or known alcohol abuse	Advanced liver fibrosis: F2-F4 score estimate on fibroscan	-	2.3%. Association with ddI/d4T found
Maggiolo [324], Italy	Prospective cohort	582	56	55% HCVAb	n/a	Receiving NVP	Grade 3 toxicity (no more information given)	ALT – 46	5.3 per 100 person years

SH=Severe hepatotoxicity; RH=relevant hepatotoxicity; HBV=hepatitis B virus; HCV=hepatitis C virus; HBsAg=Hepatitis B Surface antigen; HCVAb=hepatitis C virus antibody; HCV RNA=hepatitis C virus rubonucleicacid; AST=aspartate aminotransferase; ALT=alanine aminotransferase; LFT=liver function test; ULN=upper limit of normal; +ve=positive; pyrs=person-years; NNRTI=non nucleoside reverse transcriptase inhibitor; PI=protease inhibitor; ART=antiretroviral therapy; HAART=highly active ART; NVP=nevirapine; EFV=efavirenz; RTV=ritonavir; SQV=saquinavir; d4T=stavudine; ddI=didanosine; All AST and ALT measurements are in IU/L; * Fibrosis score calculated using the Scheuer criteria [325]

2.10 Renal Disorders

Compared to metabolic and hepatotoxic disorders, there are fewer studies that have investigated antiretroviral associated renal disorders. HIV is known to be associated with nephropathy, particularly in those of black ethnicity, and studies have reported that the administration of HAART aids in the recovery from this [326-330]. Nonetheless, there is some evidence of the occurrence of HAART-related renal disorders, which are usually found to be associated with particular antiretroviral drugs. Studies have highlighted a possible link between renal disorders and raised creatinine levels with the antiretrovirals indinavir [198-200;331-333], ritonavir [334] and tenofovir [335;336]. Table 2.3 summarises the findings of studies investigating the prevalence of renal disorders amongst those on antiretroviral regimens.

The symptoms relating to renal disorders are varied. They include renal colic, dysuria, loin pain, lithiasis, renal failure, and sterile leukocyturia. If the combination of specific symptoms used to define renal toxicity differs between studies then this will impact on the rates of renal toxicity that are observed. For example, Padberg et al [198] found an incidence of 18.8% of renal colic, whereas Dieleman et al [333] found an incidence of 24% of those with persistent sterile leukocyturia.

The most common laboratory marker used to describe renal failure is the serum creatinine level. Again, there are discrepancies in the levels used to define elevated creatinine levels. Boubaker et al [200] defined an elevated creatinine level as one that was more than 20% above the baseline value and was in the abnormal range, whereas Padberg et al [198] defined elevated creatinine levels as greater than 1.4 mg/dl. The calculated Glomular filtration rate (GFR) can also be used as a measure of renal function. This can be calculated using a number of different methods, accounting for gender, ethnicity and creatinine level, or body mass index and creatinine level [337]. Clearly, this can lead to discrepancies in incidence rates of these elevated creatinine levels as in the case of hepatotoxicity.

Table 2.3 – Summary of studies investigating the relationship between HAART and renal disorders

First author and country	Study type	n	% male	Average age	Treatment information	Primary Endpoint	Main Result
Herman[199], London	Retrospective cohort	781	n/a	n/a	All were receiving IDV for at least 1 week.	Incidence of renal complications. Creatinine levels >120 µmol/l	Renal complications: 7.3% (6.7 per 100 pyrs); 53% loin pain, 42% renal colic, 19% dysuria. Creatinine: no progressive rises
Padberg[198], Germany	Retrospective cohort	111	85	n/a	IDV (n=32), RTV+IDV (25), RTV+SQV (54)	Prevalence of renal colic, serum creatinine >1.4 mg/dl	Renal colic: 6 patients in IDV vs. 0 in the other 2 groups Creatinine: 3 (9.4%) in IDV, 3 (12.0%) in RTV+IDV, 0 in RTV+SQV
Boubaker[200], Switzerland	Retrospective cohort	106	69	40	IDV for at least 2 months	Serum creatinine >20% value at baseline into the abnormal range Cases of urinary tract complications	Creatinine: 20 (18.6%) UTI: 13 (4 lithiasis, 5 renal colic, 4 pain/dysuria)
Dieleman[333], Netherlands	Prospective cohort	184	80	39	IDV	Presence of sterile leukocyturia,	32/134 (24%) 13 of had symptoms of nephrolithiasis,
Deray[334], France	Retrospective cohort	87	n/a	n/a	RTV without SQV	Development of renal failure, Cockcroft definition of creatinine clearance	Renal Failure: 12 (13.7%) Creatinine clearance: reduced by 116 ml/min to 71 ml/min in 12 pts
Gallant et al[336], France	RCT	592	n/a	n/a	Previously naïve, receiving d4T/3TC/EFV or TFV/3TC/EFV	Development of proteinuria, glucosuria, rate of creatinine clearance at 96 weeks	Proteinuria: 21% d4T/3TC/EFV arm, 19% TFV/3TC/EFV Glucosuria: 2% d4T/3TC/EFV arm, 2% TFV/3TC/EFV arm Creatinine clearance: from 122 to 123 mL/min d4T/3TC/EFV arm, 125 to 129 mL/min TFV/3TC/EFV arm

Izzedine, France [338]	2 RCTs	(1) 189 (2) 552	n/a	n/a	(1) TFV 75mg/day, TFV 150 mg/day, TFV 300 mg/day or placebo (2) Treatment experienced, received TFV 300 mg/day or placebo	Grade 1 elevation in creatitine clearance. Grade 1 or 2 hypophosphataemia	(1) Creatinine: TFV75=6%, TFV150=4%, TFV300=2%; no grade 2 or higher elevations Phosphate: TFV75=17%, TFV150=14%, TFV300=24%; (2) Creatinine: TFV=2%, placebo 1%; no grade 2 or higher elevations Phosphate: TFV=12%, placebo=7%
Julg [339], Germany	Retrospective cohort	108	n/a	n/a	Patients receiving TFV and patients not receiving TFV but receiving ART	GFR at 12 months	Receiving TFV: 100.6 ml/min/1.73m^2; Not receiving TDF: 94.48 ml/min/1.73m^2
Wyatt [340] USA	Retrospective cohort	77694	62	44	All HIV+ve patients discharged from acute care hospitals in 1995 (pre-HAART era) and 2003 (HAART era)	Acute renal failure	More common when compared to HIV negative population Adjusted OR for HAART era=2.02 (1.86, 2.18)
El Sahly [341], USA	Retrospective cohort	488	n/a	n/a	All patients who received TFV with a pre-TFV serum creatinine and a measurement after 24 months	Percentage increase in serum creatinine: grade 1, 0–25.0%; grade 2, 25.1–50.0%; grade 3, 50.1–75.0%; grade 4, 75.1–100%; grade 5 > 100%	No change/decrease: 54% Grade 1: 33% Grade 2: 9% Grade 3: 2% Grade 4: 1% Grade 5: 1%
Jaroszewicz [342], Poland	Cross sectional study	77	71	34	Patients receiving and not receiving HAART	Mean (SD) Cystatin C, urea, creatinine, GFR	Cystatin C: HAART= 992 (63); no HAART=916 (371) ng/ml Urea: HAART= 26 (3); no HAART= 31 (2) mg/dl Creatinine: HAART= 0.71 (0.05); no HAART= 0.77 (0.03) mg/dl GFR: HAART= 138 (30); no HAART=93 (6) ml/min/1.73m^2
Gerard [343], France	Prospective cohort	53	96	41	Starting TDF/ATA/RTV+ another NRTI with multiple treatment failure	Median change at 48 weeks in creatinine clearance and serum creatinine	Creatinine clearance: +0.04 mg/dl Serum creatinine: -7.8 ml/min

Antoniou [344], Canada	Retrospective cohort	172	93	46	Received TFV for a median of 16 months	Serum creatinine>44 µmol/l; Proteinuria <3 g/l	Serum creatinine: 4% Proteinuria: 43% Discontinued TFV due to abnormal urinalysis.high creatinine: 2.3%
Jones [345], UK	Prospective cohort	4183	n/a	n/a	All HIV positive patients in clinic	Creatinine >120 µmol/l	28% ever high levels. No association with a particular ARV
Moreno [346], Spain	Prospective cohort	1286	69	40	All patients at centres in 2002-2007 receiving TFV	Discontinuation of TFV for renal dysfunction	0.4%
Buchacz [347], USA	Prospective cohort	309	86	43	Receiving TFV- or non-TFV- containing ART or naïve	Median 12 month change in Creatinine Clearance, GFR	Creatinine clearance: TFV exposed=-5.1; TFV unexposed/naïve=-2.8 mL/min p=0.51 Creatinine clearance: TFV exposed=-0.6; TFV unexposed/naïve=-0.5 mL/min/1.73m^2 p=0.55
Overton [348], USA	Cross sectional study	1694	63	40	Matched pairs of HIV positives with HIV negative controls	Mean GFR; Chronic kidney disease defined as GFR <60 mL/minute/1.73 m^2	HIV +: mean 95.5 mL/minute/1.73 m^2 4% had chronic disease; lower amongst those receiving HAART HIV +: mean 103.6 mL/minute/1.73 m^2 2% had chronic disease
Mocroft [349], Europe	Cross sectional study	4474	n/a	n/a	All under follow-up	2 consecutive GFR <60 mL/minute/1.73 m^2	Median GFR: 94.4; 3.5% with chronic renal failure; more likely to have received potentially nephrotoxic drugs and HAART
Goicoecha [350], USA	Prospective cohort	147	n/a	40	Receiving (1) TDF+PI, (2) TDF+NNRTI or (3) non-TDF containing ART	Rate of GFR decline to 48 weeks	Non-TDF: comparable to TDF+NNRTI TDF+NNRTI: decline of 6.2 mL/minute/year TDF+PI: decline of 13.9 mL/minute/year

GFR=glomular filtration rate; TFV=tenofovir; 3TC=lamivudine; ATA=atazanavir; RTV=ritonavir; SQV=saquinavir; IDV=indinavir; NRTI=nucleoside reverse transcriptase inhibitor; HAART=highly active antiretroviral therapy; OR=odds ratio

2.11 Conclusion

Whenever the incidences of HAART-related toxicities are studied it is important to interpret the results from these studies in the context of the undoubted benefits of HAART. Within this context, it is clear that the problem of antiretroviral-related toxicities must be addressed. However, the wide variation between studies in the reported incidence of these events, and their relationship with HAART often leaves readers confused as to the true impact of these events. Several factors impact on the results of studies studying this question. Differences in the way that outcome measures are defined, the demographics of the population being studied and the different antiretrovirals used in the HAART regimens will affect the prevalence rates reported. Although RCTs are the gold standard when investigating the efficacy of antiretroviral regimens, they do have the limitation of strict entry criteria and short periods of follow up. Therefore, the results from RCTs may not be generalisable to the whole HIV positive population. Although observational databases contain information on complete populations, they are, however, susceptible to biases caused by unknown confounders. These issues should be taken into consideration when investigating HAART-related toxicities, and potential solutions to some will be investigated in subsequent chapters.

Chapter 3 – Methods and Data Collection

3.1 Introduction

In this thesis, I have carried out analyses on two datasets. This chapter describes the data collection methods for the datasets upon which analyses have been undertaken. Investigations of the issues surrounding HIV toxicities in adults are conducted in the Royal Free HIV clinic database. For the analysis of total cholesterol levels in children in Chapter 8, information was obtained from the Great Ormond Street Hospital, St Mary's Hospital and the Collaborative HIV Paediatric Study (CHIPS) cohort.

3.2 The Royal Free HIV cohort

3.2.1 The Royal Free HIV clinic database

The Royal Free HIV cohort is an observational database containing information on patients that has been collected as part of routine clinical care. No additional tests or interventions are carried out for the database. It was first set up in 1994 and contains data on all HIV-positive patients who have attended the outpatients' clinic at the Ian Charleson Day Centre (ICDC) at the Royal Free Hospital, London for routine clinical care. The ICDC opened in 1991, and so a retrospective chart review was performed when the database was set up in 1994 to extract information on all patients seen between 1991-1994 by Amanda Mocroft. Thus all patients ever treated at the ICDC are included in the cohort. From 1994 to the present day, data have been collected prospectively.

At a patient's first visit a full clinical and social history is taken, which is recorded on a report form (see Appendix A) and computerised on the HIV/AIDS clinic database. Details of the previous medical history of patients who transfer to the Royal Free Hospital from other HIV clinics are requested in writing, if agreed to by the patient, and included in the database wherever possible. At subsequent routine clinic visits, the clinician completes a clinic visit form, upon which they record information on the individual's antiretroviral treatment, use of prophylaxis against *Pneumocystis carinii* pneumonia (PCP), the occurrence of any new AIDS defining illnesses and adherence to antiretroviral treatment. A copy of this form is also included in Appendix A.

Approximately every nine months, a trained research assistant, Clinton Chaloner, audits all of the data kept on the clinic database. He updates information on changes to

antiretroviral therapy that have occurred since the last notes audit, as well as information on prescription of anti-PCP treatment, dates of hospital in-patient admission and discharge, reasons for admittance as an in-patient, date and information on any new AIDS-defining events, any information on adherence noted by the clinician in the patient notes, and the date and result of any weight measurements carried out. This information is then manually entered onto the clinic database and sent to Fiona Lampe and myself. Additionally, routine laboratory data, including all measurements collected on CD4 cell count and percentage, CD8 cell count and percentage and HIV RNA viral load are transferred electronically from the appropriate departments.

We carry out checks to look for discrepancies in the database (such as date of death occurring before the date of an inpatient admission), and also to check for any potential irregularities (such as a CD4 cell count of greater than 10000 cells/mm^3, receiving more than 5 antiretroviral drugs at one time). These discrepancies are investigated and corrected where necessary. We then use all of this information to create a single SAS database (SAS Institute Inc., Cary, NC, USA), which can be used for analysis until the time of the next update. Data variables collected in the Royal Free Cohort research database are shown in Table 3.1.

As patients are still being seen at the Royal Free Hospital, the HIV Clinic database is regularly updated. However, as the audit of patient notes to extract treatment and clinical history began in October 2005 and finished in March 2006, the patients who were audited first may only have information available up until October 2005. Data downloaded from electronic sources are available until June 2006.

Table 3.1 – Data collected for the Royal Free HIV cohort

Demographics	Gender
	Ethnicity
	Date of birth
	Primary risk factor for HIV transmission
Clinical data	Date of first recorded HIV positive test
	Date of last recorded HIV negative test
	Date of first clinic visit
	Name of previous clinic attended
	Date of death
	Cause of death
	Date and description of each AIDS-defining event
	Date of admission, date of discharge and reasons for each admission to inpatient facilities
	Weight and date measured*
	Date of clinic visits*
Laboratory markers	Date and result of CD4 cell count for all available measurements
	Date and result of CD8 cell count for all available measurements
	Date and result of CD4 percentage for all available measurements
	Date and result of CD8 percentage for all available measurements
	Date, result of HIV RNA viral load and assay used for all available measurements
Antiretroviral treatment	Date of starting all antiretrovirals
	Dosage of all antiretrovirals
	Date and reason for stopping (up to 3 reasons can be given) all antiretrovirals
PCP Prophylaxis	Medications being taken at date of notes review
Adherence	Where recorded in the notes, and date information was noted

* collected from April 2003 onwards

3.2.2 Laboratory methods

The Royal Free Hospital measures T lymphocyte cell counts (including CD4 cell counts and CD8 cell counts) using standard flow cytometry techniques. Plasma HIV-1 RNA viral load has been measured by a variety of commercially available methods which have changed during the study period as improvements in assays have been made to allow capture of non-B subtypes as well as subtype B and as ultra-sensitive assays with lower limits of detection have been introduced. When first introduced in 1996, viral load monitoring was performed using the AMPLICOR PCR HIV-1 MONITOR test 1.0 (Roche Diagnostics, Roche Products Ltd., Welwyn Garden City, Hertfordshire, UK) and then later upgraded to the 1.5 version with add-in non-B primers. More recently the Cobas assay was used, which is equivalent to the Roche Assay. Prior to the introduction of a new viral load assay, the laboratory performed formal comparisons of the previous and new assay. The new assay was only accepted if results were linearly associated (R^2 >0.9) and strongly correlated (R > 0.9) [351].

3.2.3 Toxicity data

In order to investigate HAART-related toxicities, I required further information on laboratory markers that are used to define toxicities. Measurements of these laboratory markers for HIV-positive patients are carried out according to clinician judgement. Results of all laboratory tests requested at the Royal Free Hospital are stored on the hospital's pathology database by the responsible departments. Therefore, the additional data required were obtained from the Biochemistry and Haematology departments. Table 3.2 describes the information that has been obtained from these departments. It also lists the normal ranges quoted by the laboratory in 2002; these may have since changed. All available measurements for each patient seen at the ICDC were obtained. These data were first requested in July 2002. For the first download, data were collected retrospectively to get a complete record of the laboratory tests ever carried out on all patients who had ever received care at the ICDC. Firstly, results of tests that had taken place up until October 2000 were retrieved from the hospital's archived database, by searching by hospital number for each individual in the Royal Free HIV cohort database. Data from October 2000 to July 2002 were collected from the live pathology databases, and the results of all tests requested by ICDC were downloaded.

Table 3.2 – Data collected from Biochemistry and Haematology departments, and hospital-defined normal ranges in HIV-negative individuals

Laboratory marker	Department responsible for data	Normal range and units
Total cholesterol	Biochemistry	<5.2 mmol/l
HDL-cholesterol	Biochemistry	>1 mmol/l
LDL-cholesterol	Biochemistry	<4 mmol/l
Triglycerides	Biochemistry	<2.3 mmol/l
Lactate	Biochemistry	0.5-2.0 mmol/l
Glucose	Biochemistry	Fasting <5 days: 0.7-4.2 mmol/l
		Fasting >5 days: 2.9-5.3 mmol/l
Fasting status	Biochemistry	
AST	Biochemistry	5-40 IU/l
ALT	Biochemistry	5-40 IU/l
Albumin	Biochemistry	35-50 g/l
Bilirubin	Biochemistry	5-17 Umol/l
Alkaline phosphate	Biochemistry	42-128 IU/l
Gamma GT	Biochemistry	Male: 9-54 IU/l
		Female: 8-35 IU/l
Urea	Biochemistry	3.0-6.5 mmol/l
Creatinine	Biochemistry	60-97 µmol/l
White blood count	Haematology	3.7-9.5 x 10^9 cells/l
Haemoglobin	Haematology	Male: 13.5-17.5 g/dl
		Female: 11.5-15.5 g/dl
Platelets	Haematology	140-400 x 10^9 cells/l
MCV	Haematology	80-96 fL
Neutrophils	Haematology	1.7-7.5 x 10^9 cells/l
Total lymphocytes	Haematology	1.0-3.5 x 10^9 cells/l
Monocytes	Haematology	0.2-1.0 x 10^9 cells/l
Eosinophils	Haematology	0.03-0.46 x 10^9 cells/l
Basophils	Haematology	0.02-0.20 x 10^9 cells/l

HDL=high density lipoprotein; LDL=low density lipoprotein; AST=aspartate aminotransferase; ALT= alanine aminotransferase; Gamma GT= Gamma-Glutamyl Transpeptidase; MCV=mean corpsular volume

Data from the Haematology department came in the form of comma separated value (CSV) files, and data from the Biochemistry department came in the form of fixed field length files. I converted these to SAS files and then carried out checks to look for any obvious discrepancies. The date of all measurements taken was compared to the date of first visit to the ICDC, and to the date of death to ensure all measurements occurred within this window; duplicate measurements are deleted, and the values of the laboratory markers themselves are checked to ensure that none were negative or implausible. From July 2002 onwards (when the first data download occurred), updates of the toxicity data have occurred approximately every 9 months, to coincide with the updates of the main Royal Free HIV clinic database.

Clinicians are able to request several laboratory tests on the same blood sample. Therefore, batches of tests are often carried out at the same time. Although clinicians do not always order all tests, frequently they will request a group of tests, and so the dates of laboratory tests are often identical. This is discussed further in Chapter 4.

3.2.4 Data on Hepatitis B and C virus status

As noted in Chapter 2, the hepatitis B and hepatitis C status of individuals can also impact on the occurrence of HAART-related toxicities. Therefore, I also first obtained information on the hepatitis B virus and hepatitis C virus status of individuals in the cohort from the Virology department in July 2002. Similarly to the data obtained from the haematology and biochemistry departments, data on all tests carried out before September 2001 were downloaded from the hospital's archived database according to patient hospital number, and data from this time point forward were obtained from the hospital's pathology database. However, as information on hepatitis B surface antigen prior to September 2001 was found to be unreliable, results for this test prior to this date have not been used. Information was transferred in CSV format. I carried out a number of checks for discrepancies and implausible values in a similar way to those carried out on data obtained from the biochemistry and haematology departments, and converted the files to SAS files so that they could be used in analyses. Data have been collected prospectively from July 2002 onwards. Information on the measurements obtained from virology, and the form that the results of the test take are shown in Table 3.3.

Table 3.3 – Data collected from Virology department to assess hepatitis B and hepatitis C virus status

Marker	Abbreviation	Data collected	Results possible	Units
Hepatitis B surface antigen	HBsAg	Detection of HBsAg	Positive, Negative/Equivocal, Weak positive	-
HBV DNA viral load		Detection of HBV DNA	Levels below lower limit of detection, measurable viral load, viral load level above threshold of assay	-
		Level of HBV DNA detected	Value of viral load or lower/upper limit of assay	Copies/ml, Megagenome equivalents/ml, IU/ml or IU/L
Hepatitis B core antibody	HBcAb	Detection of HBcAb	Positive, Negative/Equivocal, Weak positive	-
Hepatitis B surface antibody	HBsAb	Detection of HBsAb	Positive, Negative/Equivocal, Weak positive	-
		Detection of HBsAb	Levels below lower limit of detection, measurable viral load, viral load level above threshold of assay	-
		Level of HBsAb detected	Value of viral load or lower/upper limit of assay	IU/L
Hepatitis C virus antibody	HCVAb	Detection of HCVAb	Positive/reactive, Negative/Equivocal, Weak positive	-
HCV RNA viral load		Detection of HCV RNA	Levels below lower limit of detection, measurable viral load, viral load level above threshold of assay	-
		Level of HCV RNA detected	Value of viral load or lower/upper limit of assay	Megagenome equivalents/ml
HCV Genotype		Genotype	1, 1A, 1B, 2A, 2C,3 ,3A ,4 ,4C ,4D	-

3.3 Great Ormond Street and St Mary's Hospitals and the Collaborative HIV Paediatric Study (CHIPS) Cohort

3.3.1 Introduction

The data from Great Ormond Street (GOS) Hospital, St Mary's Hospital and the CHIPS database were collected specifically for a project investigating the association between antiretroviral treatment and total cholesterol measurements. I shall describe in this section the data that were collected and how the data from different sources were merged.

3.3.2 Total cholesterol data from St Mary's and GOS Hospitals

Great Ormond Street Hospital, London and St Mary's Hospital, London both have specialist clinics for HIV-positive children and adolescents. As part of routine care, non-fasting total cholesterol measurements are performed on children. For this study we obtained information on all total cholesterol measurements from every clinic appointment from September 1995 until January 2006 for each child aged 16 years or younger from the two hospitals. Ethical approval was obtained from the committees of the Institute of Child Health and Great Ormond Street Hospital NHS trust and St Mary's NHS Trust. The research was conducted in accordance with guidelines for human experimentation as specified by the NHS Trust Ethical Committees for St Mary's and Great Ormond Street Hospitals.

3.3.3 The Collaborative HIV Paediatric Study (CHIPS) Cohort

The Collaborative HIV Paediatric Study (CHIPS) Cohort was established in April 2000 and is a multi-centre cohort study of HIV-1 infected children in the United Kingdom and Republic of Ireland [352]. Initially, 16 UK and Irish centres participating in the Paediatric European Network for the Treatment of Aids (PENTA) trials coordinated at Medical Research Council Clinical Trials Unit (MRC CTU) were included in CHIPS. There are now 43 British and Irish centres (including GOS Hospital and St Mary's Hospital) caring for HIV infected children who collaborate in the CHIPS cohort, along with the National Study of HIV in Pregnancy and Childhood (NSHPC) and the MRC CTU. Information collected includes growth data, clinical events and CDC stage of disease, results of T-cell subsets and HIV RNA viral load tests, hospital admissions and outpatient visits, antiretroviral therapy (including doses prescribed, start and stop dates, reason for change, adverse events) and information on lipodystrophy and puberty development. Limited total cholesterol measurements are also available when supplied by the

hospital. However, only one measurement every quarter per child is recorded in the CHIPS database, and thus these data were expected to contain many fewer observations than that obtained from the individual hospitals. More information on the CHIPS cohort is available at the following website: http://www.bhiva.org/chiva/protocols/chips.html.

For this particular project, data on antiretroviral treatment history, CD4 count, viral load and demographics of patients attending St Mary's and GOS Hospitals were obtained from the CHIPS database. Additionally, I also obtained any total cholesterol measurements recorded on the CHIPS database for these children. It was expected that many of these measurements would duplicate those obtained from the direct downloads from GOS and St Mary's. However, other measurements on the children (for example those carried out at other hospitals if the child transferred their care) may also be available. The CHIPS database retains each child's hospital number, as well as a unique CHIPS identifier, and hospital numbers were used to link data from the CHIPS database with total cholesterol information supplied by GOS and St Mary's. Data were available until March 2006.

3.3.4 Data amalgamation and cleaning

In total, 184 children were identified by GOS and 149 children were identified by St Mary's as ever having a total cholesterol measurement. Of the 184 children identified by GOS, 18 did not match with patients in the CHIPS database and were removed from analyses for risk of duplication. Of the 149 children identified by St Mary's, 17 did not match with patients in the CHIPS database and were also removed from analyses. Therefore, 298 patients with total cholesterol measurements were identified by the two hospitals.

Three hundred and seventy two children were identified in the CHIPS database as having attended at least one of the two hospitals for care, and as having at least one total cholesterol measurement. Eighty-seven children had measurements only available from the CHIPS database (potentially measured at other hospitals), and 285 children had measurements reported from both sources. Additionally, 13 children only had total cholesterol measurements from the data supplied by GOS and St Mary's (but demographic and treatment data for these children was available from CHIPS). Therefore 385 children could be included in analyses.

A total of 5568 lipid measurements were available from the three sources. Of these, 1155 (20.7%) were found in both the GOS/St Mary's database and the CHIPS

database, 4009 (72.0%) measurements were only in the GOS/St Mary's database and 404 (7.3%) were only available in the CHIPS database. The 404 measurements on 113 children that are only available from the CHIPS database may have been taken at other hospitals. They were checked visually, to ensure that they were consistent with other measurements taken and, as this appeared the case, they were included in analyses.

Repeated samples on the same day were observed in 7 cases. There were 2 samples on the same day in 6 cases, and 3 samples on the same day in 1 case. As the maximum difference between values measured on the same day was 0.25 mmol/l, these samples were averaged to get a single value per day for each child. Therefore, there were a total of 5560 total cholesterol observations on 385 children available for analyses.

3.4 Statistical Methods

As the focus of this thesis is to investigate the methodological issues associated with investigating antiretroviral-related toxicities, many of the statistical methods used are described in detail in the appropriate chapter. Nonetheless, there are also standard statistical methods that I have used throughout the thesis. Regression models have been used for situations in which there is only one outcome measurement per individual included in the analysis – linear regression for continuous outcome measures and logistic regression for binary outcomes. When investigating the time taken to experience an event (e.g. death, occurrence of a toxicity), I have used Cox proportional hazards models. For the situation in which we have repeated observations on individuals at more than one time point I have used mixed effects models (also known as multi-level models). The basic statistical models underlying these methods are described in detail in Appendix B.

3.5 Conclusions

This chapter has summarised the information collected in the Royal Free HIV Clinic database, and from Great Ormond Street Hospital, St Mary's Hospital and the CHIPS cohort. These datasets are used in subsequent chapters to investigate potential biases when assessing the prevalence and incidence of antiretroviral-related toxicities.

Chapter 4 – The impact of calendar time on the frequency of monitoring and the use of HAART in the Royal Free Cohort

4.1 Introduction

In Chapter 2, I considered the potential biases that may occur when assessing HAART-related toxicities. One such bias may be introduced if the frequency of monitoring of laboratory markers used to define the occurrence of antiretroviral-related toxicities has changed over time, or varies by demographic group or among those receiving different antiretroviral regimens. In this chapter I shall investigate this issue further. I shall start by investigating whether there is any evidence that this bias is present in the Royal Free cohort amongst previously antiretroviral-naïve patients starting HAART for the first time. These analyses will enable me to identify whether differential frequency of monitoring in different calendar years has occurred, and thus whether this is likely to introduce bias when comparing the prevalence of HAART-associated toxicities amongst different population groups or among those receiving different antiretrovirals. Subsequently, I shall investigate the impact of this bias on the estimated prevalence of antiretroviral toxicities and the analytical methods least affected by these biases. I wish to consider simple endpoints that could be used in every day analyses, rather than more complicated analytical techniques. To achieve these aims, I will perform data simulations to identify the methods for comparing the prevalence of HAART-related toxicities that are least affected by changes in the frequency of monitoring. This will provide some insight into the most appropriate endpoint to use when considering toxicities.

4.2 Methods

4.2.1 Changes over time in the characteristics of those starting HAART in the Royal Free Cohort

For analyses in this chapter, I included all previously antiretroviral naïve patients at the Royal Free Hospital who started HAART from 1st January 1998 until 31st December 2004. This ensured that individuals were starting HAART from the time of widespread use (i.e. patients in early RCTs of HAART were excluded), and also gave the potential for at least one year of follow-up for each person. Furthermore, prior to 1998, the frequency of monitoring of certain laboratory markers, such as HDL cholesterol, was so

low that these early years would have to be excluded from any analyses of antiretroviral-related toxicities. Individuals were also required to have a CD4 cell count measurement and viral load measurement in the six month period prior to starting HAART to ensure that individuals were under regular follow-up when starting HAART. HAART was defined as an antiretroviral regimen of at least three antiretrovirals including one of a PI, an NNRTI or abacavir.

I first considered the frequency of monitoring of the laboratory markers used to measure, at least in part, any evidence of drug-related toxicity. Laboratory tests ordered from the biochemistry department are often ordered in batches, in the groupings shown below:

1 **ALT**, creatinine, albumin, alkaline phosphate, AST, bilirubin, gamma GT and urea
2 **Total cholesterol** and triglycerides
3 **Glucose**
4 **HDL cholesterol**, LDL cholesterol

Therefore, I have chosen to focus on describing the presence of four laboratory markers in particular, one chosen from each group: ALT, total cholesterol, glucose and HDL cholesterol. For each laboratory marker and each individual, I assessed whether a pre-HAART measurement was taken, the number of measurements in the first year of HAART, the time to the first measurement after starting HAART, and the average (mean) time between measurements in the first year of HAART. I also calculated whether the individual had a measurement twelve months after starting HAART. This was defined as the presence of a measurement in the period 10 to 14 months after starting HAART. The results were then summarised for all individuals, stratified by the year of starting HAART.

If any other factors that have also changed over time, such as demographic factors and the antiretrovirals prescribed, then any analysis of the impact of these factors on the occurrence of HAART-related toxicities will also be affected by differential frequency of monitoring and may give biased results. Therefore, I next summarised the characteristics and antiretrovirals received according to the year of starting HAART to investigate whether this was likely to be an important issue.

Finally, I considered the factors associated with the presence of a pre-HAART laboratory measurement and a measurement one year after starting HAART for each of the four laboratory markers. Potential explanatory factors considered were year of HAART (categorical variable), CD4 count when starting HAART (per 100 cells/mm^3 higher), viral load when starting HAART (per 1 log copies/ml higher), age at HAART (per 10 years older), risk for HIV transmission (homosexual, heterosexual or other), ethnicity (white, black African or other), gender (male or female), HBV status (positive, negative or unknown), HCV status (positive, negative or unknown), and the NNRTI/PI included in the regimen (EFV only, NVP only, IDV only, NFV only, LPV/r only, ABA only or other). Patients were considered to have HBV if at any time they had a surface antigen test (HBsAg) which was positive or weakly positive, regardless of the timing of the test (i.e. whether it occurred prior to or after starting HAART). Similarly, patients were considered to have HCV if at any time they had a positive or weakly positive core antibody test (HCVAb). For the analysis considering factors associated with a one-year measurement, an additional potential explanatory factor was the presence of a pre-HAART measurement (missing, abnormal [defined based on the Royal Free laboratory normal range for the marker] or normal). Analyses were carried out using logistic regression models (see Appendix B). All potential explanatory factors were included in both univariable and multivariable analysis.

4.2.2 Creation of simulated dataset used in analyses investigating the analytical approach least affected by biases as a result of differential frequency of monitoring

Having investigated whether there was any evidence of differential frequency of monitoring, I next went on to consider the implications of this by using data simulations. This enabled me to pre-define whether there is a 'true' effect of receiving a particular antiretroviral regimen on the occurrence of toxicity and therefore investigate the extent to which different analytical approaches are affected by this bias.

I created a simulated dataset of 1000 hypothetical antiretroviral-naive individuals who I randomly assigned with equal probability to receive one of two HAART regimens, regimen A and regimen B (although the approach for comparing demographic factors is analogous). I wished to consider whether these regimens were associated with different occurrences of HAART-related toxicity in the first year of HAART. The actual definition of an antiretroviral-related toxicity shall remain unspecified (this is discussed further in Chapter 5), but I assumed that it was classified based on the occurrence of an abnormal value for a laboratory marker and could, for example, be the occurrence

of a high total cholesterol level or a high ALT/AST level. Thus, individuals were required to have a measurement taken in order to be defined as experiencing a toxicity. I also assumed that the occurrence of toxicity did not depend on any other factor (gender, ethnicity, etc).

I have assumed that the time to a toxicity occurring after starting HAART for the first time was exponentially distributed with a pre-specified mean (X, where I changed the value of X in different simulations). See Appendix B for a description of the exponential distribution. Furthermore, toxicities may resolve without any clinician intervention. Thus, from the date of a toxicity occurring, the time to experiencing a resolution of this toxicity was also assumed to be exponentially distributed, but this time with a pre-specified mean of Y weeks. Initially, I assumed that the rates of toxicity and toxicity resolution were the same for those receiving regimen A and for those receiving regimen B, but this assumption was modified subsequently.

I next considered how frequently the laboratory marker of interest was measured for those on regimen A and for those on regimen B. I assumed that regimen A was generally started in earlier calendar years than regimen B, and thus there was more frequent measuring of laboratory markers for those receiving regimen B. I have assumed that the time between measurements followed a log-normal distribution (i.e. the natural log of the time between measurements followed a normal distribution), as investigations suggested that this was the most appropriate (see Appendix B for definition of the normal distribution). Patients on regimen A were assumed to have their first toxicity marker measurement on average (as expressed by the geometric mean) 6 weeks after starting HAART, and measurements were subsequently every 10 weeks on average. Those receiving regimen B were assumed to have their first toxicity marker measurement on average 4 weeks after starting HAART, and subsequent measurements were taken every 7 weeks. This approximates the frequency of measuring in the Royal Free cohort described in Table 4.2.

I next assessed the range of values for the estimate of the treatment effect that would be obtained from models in which the true time at which toxicity occurred and the true time of toxicity resolution for each individual was known, and thus no bias as a result of differential frequency of monitoring existed. Whilst this is unlikely to happen in clinical practice (in the absence of continual monitoring), this gave me a reference for comparison when considering the situation in which bias due to differential frequency of monitoring existed. I considered two endpoints:

(i) The time to the occurrence of a toxicity event. An estimate of the effect of receiving regimen B compared to regimen A was obtained with the hazard ratio.

(ii) The percentage of individuals who were experiencing an event at one year (i.e. an event had occurred at this time and not yet resolved). An estimate of the effect of receiving regimen B compared to regimen A was obtained with the odds ratio.

4.2.3 Data simulations investigating the analytical approach least affected by biases as a result of differential frequency of monitoring – no true treatment effect

I have therefore defined the assumptions required for my initial data simulations. I have created a hypothetical dataset of patients for whom I know the 'true' impact of receiving regimen B compared to receiving regimen A on the occurrence of antiretroviral-related toxicities, (i.e. there is no effect as it is assumed to be the same for both treatment groups), and I have a reference range with regard to the parameter estimates likely to be obtained. I also know that those on treatment regimen A have fewer laboratory marker measurements during the first year of HAART. Therefore, I can now choose a number of endpoints that I might consider using in any real life study to investigate the occurrence of antiretroviral related toxicities, and can identify the which give least biased results (i.e. they correctly find, on average, that there is no difference between the two treatment regimens with regard to the occurrence of toxicities). The endpoints I have chosen to investigate are:

(1) The median *time* to experiencing a toxicity event

(2) The *proportion* having at least one measurement in the first year of HAART that meets the definition of a toxicity event. Patients who do not have a value in this window were assumed to have experienced an event (missing=failure)

(3) The *proportion* having at least one measurement in the first year of HAART that meets the definition of a toxicity event. Patients who do not have a value in this window were excluded (missing=excluded)

(4) The *proportion* whose first measurement in the period six months to one year after starting HAART meets the definition of a toxicity event. Patients who do not have a value in this window were assumed to have experienced an event

(5) The *proportion* whose first measurement in the period six months to one year after starting HAART meets the definition of a toxicity event. Patients who do not have a value in this window were excluded
(6) The *proportion* whose measurement at one year (the measurement in the window 10-14 months after starting HAART that occur closest to one year) meets the definition of a toxicity event. Patients who do not have a value in this window were assumed to have experienced an event
(7) The *proportion* whose measurement at one year after starting HAART meets the definition of a toxicity event. Patients who do not have a value in this window were excluded

I then repeated this simulation 1000 times. For each simulation, and for each endpoint described above, I calculated the hazard ratio (HR) for endpoint 1 and the odds ratio (OR) for endpoints 2 to 7. In all situations I compared those receiving regimen B to those receiving regimen A. I also noted for each simulation whether the p-value assessing whether there was a difference between the two different treatment regimens was less than 0.05 (using a log-rank test for endpoint 1, and a chi-squared test for all other endpoints). If there were no biases in our cohort, we would expect to observe a p-value of less than 0.05 in 5% of simulations. We would also expect the mean log odds ratio/log hazard ratio to equal zero, and it to be greater than zero in approximately 50% of occasions.

I first began by assuming that X, the mean time to the occurrence of a toxicity, was equal to 26 weeks, and that Y, the mean time to resolution of the toxicity, was equal to 500 weeks, as this appeared to reasonably approximate the frequency of metabolic disorders as seen in Chapter 2. I recorded the results of my data simulations under these assumptions. I then varied the values for X and Y to investigate impact of the choice of these values on the results obtained. At first, I continued to assume that the time to occurrence of a toxicity, and to resolution of toxicity was the same for those on regimen A and on regimen B (i.e. there was no treatment effect). In all situations, I assumed that the time to the occurrence of a toxicity and the time to resolution of the toxicity followed exponential distributions. I considered situations 1 to 8 described in Table 4.1.

Table 4.1 – Geometric mean time to occurrence of a toxicity and to toxicity resolution assumed for each data simulation situation

	Regimen A		Regimen B	
	Mean time to toxicity event (weeks)	Mean time to toxicity resolution (weeks)	Mean time to toxicity event (weeks)	Mean time to toxicity resolution (weeks)
	Same incidence of toxicity in both groups			
(1)	26	500	26	500
(2)	52	500	52	500
(3)	200	500	200	500
(4)	500	500	500	500
(5)	26	200	26	200
(6)	52	200	52	200
(7)	200	200	200	200
(8)	500	200	500	200
	Incidence of toxicity is higher in amongst those receiving regimen A			
(9)	26	500	52	500
(10)	26	500	39	500
(11)	26	500	30	500
(12)	200	500	400	500
(13)	200	500	300	500
(14)	200	500	250	500
	Incidence of toxicity is higher in amongst those receiving regimen B			
(15)	52	500	26	500
(16)	39	500	26	500
(17)	30	500	26	500
(18)	400	500	200	500
(19)	300	500	200	500
(20)	250	500	200	500

4.2.4 Data simulations investigating the analytical approach least affected by biases as a result of differential frequency of monitoring – other situations considered

I next investigated the results that would be obtained in the situation where there was a true difference between the two treatment groups. I firstly assumed that the time to the occurrence of a toxicity amongst those on regimen A was shorter than that amongst those on regimen B, i.e. there was a higher incidence of toxicity amongst those receiving regimen A. The circumstances considered are described in Table 4.1 (situations 9 to 14). I then assumed the opposite: that the incidence of toxicity was higher amongst those on regimen B than amongst those receiving regimen A. The specific circumstances considered are described in Table 4.1 (situations 15 to 20).

In all situations considered so far, the only factor that could affect the frequency of monitoring is the treatment regimen received (i.e. all other factors are equal in the two groups). However, this may not always be the case – if a patient is feeling unwell as a

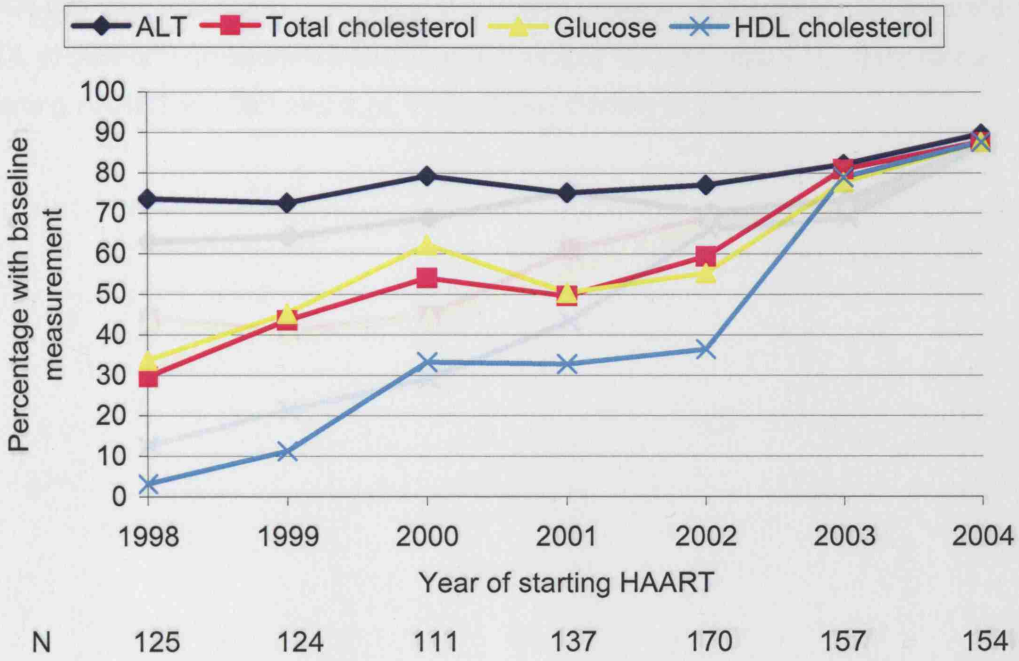
result of the toxicity they are experiencing then they may be more likely to visit the hospital and have a laboratory measurement taken. Therefore, I considered again situations 1, 3, 9, 12, 15 and 18, but added the condition that the frequency of monitoring was also dependent both on the regimen received, and on whether or not a toxicity occurred in the first year of HAART or not. As described previously, I assumed the time between measurements followed a log-normal distribution. The geometric means assumed were:

(i) Patients receiving regimen A who do not experience a toxicity in the first year of HAART:
    a. time to first measurement: 6 weeks,
    b. time between subsequent measurements: 10 weeks
(ii) Patients receiving regimen A who experience a toxicity in the first year of HAART:
    a. time to first measurement: 6 weeks,
    b. time between subsequent measurements: 8 weeks
(iii) Patients receiving regimen A who do not experience a toxicity in the first year of HAART:
    a. time to first measurement: 4 weeks,
    b. time between subsequent measurements: 7 weeks
(iv) Patients receiving treatment regimen A who experience a toxicity in the first year of HAART:
    a. time to first measurement: 4 weeks,
    b. time between subsequent measurements: 5 weeks

4.3 Changes over time in the presence of laboratory markers amongst those starting HAART in the Royal Free Cohort

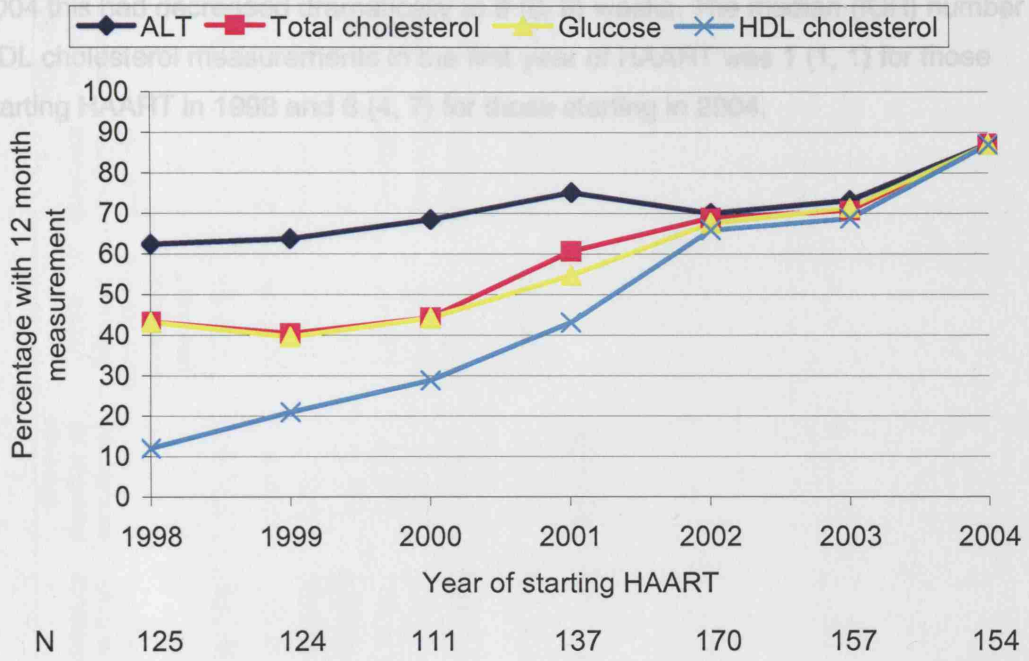
Overall, 978 previously antiretroviral-naïve individuals started HAART at the Royal Free Hospital between 1998 and 2004. The monitoring of laboratory markers has changed dramatically over time (Figure 4.1). The percentage of individuals with a baseline ALT measurement has increased from 73.6% of those starting HAART in 1998 to 89.6% of those starting HHART in 2004. There have been even more dramatic increases for the other markers: the percentage with a baseline HDL measurement has increased from 3.2% in 1998 to 87.7% in 2004.

Figure 4.1 – Percentage of patients starting HAART who have a baseline laboratory marker measurement according to the year of starting HAART



Similarly, there have been dramatic increases in the percentage with a measurement 12 months after starting HAART according to the year of starting HAART (Figure 4.2). Of those starting HAART in 1998, 62.4%, 43.2%, 43.2% and 12.0% had a 12 month ALT, total cholesterol, glucose and HDL cholesterol measurement, respectively. By 2004 these percentages had reached 87.0% for all four measurements. Thus, although the frequency of monitoring is likely to impact on the results of any analysis of these surrogate laboratory markers, the analyses involving HDL cholesterol, total cholesterol and glucose and other markers of this group are likely to be more affected than analyses of ALT levels.

Figure 4.2 – Percentage of those starting HAART with 12 month laboratory marker measurements according to year of starting HAART



When considering those with at least one ALT measurement in the first year after starting HAART (Table 4.2), the median number of observations in the first year of HAART and the average time between observations remained constant over time at around 6 measurements and 7 weeks respectively. In contrast, when considering total cholesterol measurements, the median (IQR) time between consecutive observations has decreased in more recent calendar years, from a median (IQR) of 13 (8, 23) weeks amongst those starting HAART in 1998 to 7 (6, 9) weeks for those starting HAART in 2004. The median (IQR) number of total cholesterol measurements in the first year of HAART was 3 (1, 4) for those starting HAART in 1998 and 6 (5, 7) for those starting in 2004.

Again, amongst those with at least one glucose measurement in the first year of HAART, the average time between subsequent observations has also decreased in more recent calendar years. There was a median (IQR) of 13 (8, 23) weeks between glucose measurements in the first year for those starting HAART in 1998. For those starting HAART in 2004, this interval dropped to 7 (6, 9) weeks. The median (IQR) number of glucose measurements in the first year of HAART was 2 (1, 4) for those starting HAART in 1998 and 6 (5, 7) for those starting in 2004.

The most dramatic changes over time are seen when considering HDL cholesterol levels. The median (IQR) time between subsequent HDL cholesterol measurements in the first year of HAART was 45 (24, 49) weeks for those starting HAART in 1998. By 2004 this had decreased dramatically to 8 (6, 9) weeks. The median (IQR) number of HDL cholesterol measurements in the first year of HAART was 1 (1, 1) for those starting HAART in 1998 and 6 (4, 7) for those starting in 2004.

Table 4.2 – Frequency of monitoring of laboratory measurements in the first year of HAART

Year of starting HAART	Number	ALT			Total cholesterol			Glucose			HDL cholesterol		
		Measure in first year (%)	Number of measures*	Time between consecutive measures (weeks) *	Measure in first year (%)	Number of measures*	Time between consecutive measures (weeks) *	Measure in first year (%)	Number of measures*	Time between consecutive measures (weeks) *	Measure in first year (%)	Number of measures*	Time between consecutive measures (weeks) *
1998	125	97 (78)	5 (3, 8)	8 (6, 12)	87 (70)	3 (1, 4)	13 (8, 23)	82 (66)	2 (1, 4)	13 (8, 23)	15 (12)	1 (1, 1)	45 (24, 49)
1999	124	100 (81)	7 (4, 10)	6 (4, 9)	89 (72)	3 (2, 4)	12 (7, 17)	89 (75)	3 (2, 5)	10 (6, 16)	34 (27)	2 (1, 3)	16 (7, 37)
2000	111	94 (87)	7 (4, 10)	6 (4, 8)	74 (67)	5 (2, 7)	7 (6, 10)	82 (74)	5 (2, 7)	7 (5, 9)	53 (48)	5 (1, 6)	8 (6, 11)
2001	137	120 (88)	6 (4, 8)	7 (6, 10)	112 (82)	4 (2, 6)	9 (6, 20)	105 (77)	4 (1, 7)	9 (7, 19)	88 (64)	3 (1, 5)	11 (7, 26)
2002	170	147 (87)	5 (3, 8)	8 (5, 11)	145 (85)	5 (3, 6)	9 (7, 13)	145 (85)	4 (2, 6)	9 (6, 16)	141 (83)	4 (2, 6)	11 (7, 20)
2003	157	144 (92)	6 (4, 8)	7 (6, 9)	142 (90)	6 (4, 7)	8 (6, 10)	143 (91)	5 (4, 7)	8 (6, 10)	141 (90)	5 (4, 7)	8 (6, 10)
2004	154	151 (98)	6 (5, 8)	7 (5, 9)	151 (98)	6 (5, 7)	7 (6, 9)	150 (97)	6 (5, 7)	7 (6, 9)	150 (97)	6 (4, 7)	8 (6, 9)
Total	**978**	**853 (87)**	**6 (4, 8)**	**7 (5, 10)**	**800 (82)**	**5 (2, 6)**	**8 (6, 12)**	**796 (81)**	**5 (2, 7)**	**8 (6, 13)**	**622 (64)**	**5 (2, 6)**	**8 (6, 13)**

* Median (inter-quartile range) amongst those with at least one measurement in the first year of HAART

4.4 Changes over time in the use of highly active antiretroviral therapy use and demographic factors in the Royal Free Cohort

4.4.1 Characteristics of those starting HAART in the Royal Free Cohort

Of the 978 previously antiretroviral-naïve individuals who started HAART from 1998 onwards, 719 (73.5%) were male, 517 (52.8%) had a homosexual risk for infection, 554 (59.7%) were of white ethnicity and 292 (29.9%) were of black African ethnicity (Table 4.3). The median CD4 cell count at the time of starting HAART was 189 cells/mm^3 (IQR 77 to 295 cells/mm^3) and the median viral load was 5.2 $\log_{10}$ copies/ml (IQR 4.7 to 5.7 $\log_{10}$ copies/ml).

The proportion of those starting HAART at the Royal Free who are men has steadily decreased over time, from 77.6% in 1998 to 72.7% in 2004. In contrast, the proportion that are heterosexual has increased from 29.6% in 1998 to 41.6% in 2004, and the proportion who are of Black African ethnicity has increased from 23.2% in 1998 to 31.2% in 2004. This reflects the changes in the demographics of the HIV positive clinic population at the Royal Free Hospital as a whole [353]. The median viral load at which HAART was started has remained constant over time periods. However, the median CD4 cell count at the time of starting HAART has changed. In 1998, the median (IQR) CD4 cell count for starting HAART was 216 (95, 360) cells/mm^3. By 2000 this had fallen to 147 (39, 225) cells/mm^3, and from this point has risen again to 202 (86, 285) cells/mm^3 in 2004. This reflects changes in published guidelines, and evidence from the literature concerning the appropriate time to start HAART [354-356]. The average age at the start of HAART has remained constant at around 35 years throughout the study period.

The proportion with unknown HCV status has remained constant at about 10% during the study period. The proportion that is known to be HCV positive has fallen from 15.2% in 1998 to 5.2% in 2004. In contrast, the proportion with unknown HBV status has fallen from 29.6% in 1998 to 11.0% in 2004 (although it must be remembered that information on HBV status prior to 2001 is not available). The proportion known to be infected with HBV virus has increased from 4.0% in 1998 to 7.8% in 2004.

Table 4.3 – Characteristics of previously antiretroviral-naïve individuals starting HAART from 1998 onwards in the Royal Free Cohort according to year of starting HAART

Year of starting HAART		**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**Total**
n		**125**	**124**	**111**	**137**	**170**	**157**	**154**	**978**
Gender	*Male*	97 (77.6)	94 (75.8)	81 (73.0)	96 (70.1)	128 (75.3)	111 (70.7)	112 (72.7)	**719 (73.5)**
Risk group	*Homosexual*	81 (64.8)	68 (54.8)	57 (51.4)	64 (46.7)	88 (51.8)	80 (51.0)	86 (55.8)	**517 (52.8)**
	Heterosexual	37 (29.6)	49 (39.5)	47 (42.3)	66 (48.1)	74 (43.5)	73 (46.5)	64 (41.6)	**417 (42.6)**
	Other	7 (5.6)	7 (5.7)	7 (6.3)	7 (5.1)	8 (4.7)	4 (2.6)	4 (2.6)	**44 (4.5)**
Ethnicity	*White*	87 (69.6)	78 (62.9)	58 (52.3)	74 (54.0)	102 (60.0)	73 (46.5)	82 (53.3)	**554 (59.7)**
	Black African	29 (23.2)	28 (22.6)	39 (35.1)	42 (30.7)	44 (25.9)	62 (39.5)	48 (31.2)	**292 (29.9)**
	Other	9 (7.2)	18 (14.5)	14 (12.6)	21 (15.3)	24 (14.1)	22 (14.0)	24 (15.6)	**132 (13.5)**
HCV status	*Positive*	19 (15.2)	12 (9.7)	4 (3.6)	10 (7.3)	14 (8.2)	14 (8.9)	8 (5.2)	**81 (8.3)**
	Negative	89 (71.2)	100 (80.6)	96 (86.5)	107 (78.1)	138 (81.2)	115 (73.2)	132 (85.7)	**783 (80.1)**
	Unknown	17 (13.6)	12 (9.7)	11 (9.9)	20 (14.6)	18 (10.6)	28 (17.8)	14 (9.1)	**114 (11.7)**
HBV Status	*Positive*	5 (4.0)	6 (4.8)	6 (5.4)	5 (3.7)	9 (5.3)	10 (6.4)	12 (7.8)	**53 (5.4)**
	Negative	83 (66.4)	82 (66.1)	72 (64.9)	105 (76.6)	139 (81.8)	119 (75.8)	125 (81.2)	**725 (74.1)**
	Unknown	37 (29.6)	36 (29.0)	33 (29.7)	27 (19.7)	22 (12.9)	28 (17.8)	17 (11.0)	**200 (20.5)**
Viral load (log copies/ml)	*Median (IQR)*	5.3 (5.0, 5.7)	5.5 (4.9, 5.8)	5.3 (4.9, 5.8)	5.2 (4.7, 5.6)	5.1 (4.6, 5.6)	5.1 (4.6, 5.6)	5.0 (4.6, 5.5)	**5.2 (4.7, 5.7)**
CD4 count (cells/mm^3)	*Median (IQR)*	216 (95, 360)	153 (71, 293)	147 (39, 225)	181 (75, 295)	183 (72, 299)	218 (109, 310)	202 (86, 285)	**189 (77, 295)**
Age (years)	*Median (IQR)*	34 (30, 39)	36 (32, 41)	36 (32, 41)	36 (31, 43)	36 (32, 42)	36 (32, 4.2)	37 (32, 42)	**36 (32, 42)**

Entries are number (percentage) unless otherwise stated

4.4.2 Antiretroviral regimens amongst those starting first-line HAART regimens in the Royal Free Cohort

The most common regimen type prescribed over the study period was 1 NNRTI with 2 NRTI (490/978; 50.1%), followed by a PI with RTV and 2 NRTIs (259; 26.5%) and one PI with 2 NRTIs (129; 13.2%; Table 4.4). There was great variation in the types of regimen prescribed, according to the calendar year of starting HAART. The percentage whose initial regimen included 1 PI and 2 NRTI has fell from 65.6% in 1998 to 3.3% in 2004. Conversely, the percentage starting 1 NNRTI and 2 NRTI varied from 24.8% to 51.3% between 1998 and 2004. Furthermore, there has been an increase in the percentage receiving PI+RTV and 2 NRTI as a first-line regimen, from 3.2% in 1998 to 42.2% in 2004, reflecting increasing awareness of the efficacy for ritonavir-boosted regimens [357;358]

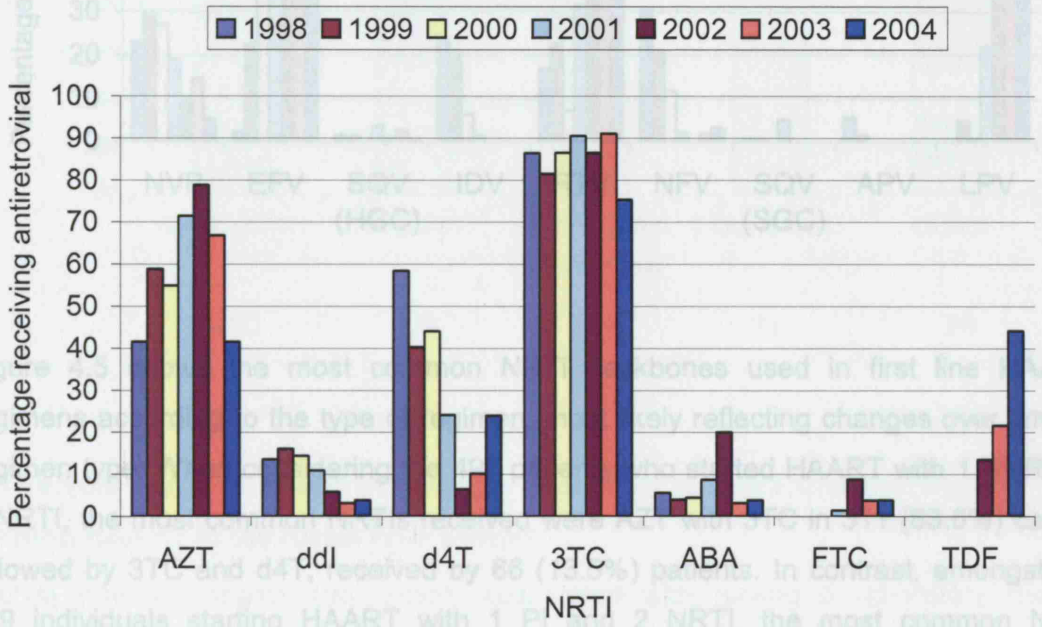
Table 4.4 – Starting regimens used in the Royal Free Cohort from 1998 to 2004

Starting regimen	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**Total**
NNRTI + 2 NRTI	31 (24.8)	62 (50.0)	81 (73.0)	77 (56.2)	68 (40.0)	92 (58.6)	79 (51.3)	490 (50.1)
PI + 2 NRTI	82 (65.6)	27 (21.8)	10 (9.0)	3 (2.2)	0 (0.0)	2 (1.3)	5 (3.3)	129 (13.2)
PI+RTV + 2 NRTI	4 (3.2)	28 (22.6)	8 (7.2)	41 (29.9)	63 (37.1)	50 (31.8)	65 (42.2)	259 (26.5)
ABA + 2 NRTI	0 (0.0)	1 (0.8)	3 (2.7)	4 (2.9)	11 (6.5)	2 (1.3)	0 (0.0)	21 (2.2)
Other	8 (6.4)	6 (4.8)	9 (8.1)	12 (8.8)	28 (16.5)	11 (7.0)	5 (3.3)	79 (8.1)
Total	**125**	**124**	**111**	**137**	**170**	**157**	**154**	**978**

Figure 4.3 shows the individual NRTIs that were used in the initial first-line HAART regimens. The most popular NRTIs overall were 3TC, which most (835; 58.4) patients received; AZT, which was in the initial regimen of 587 (60.0%) individuals, and d4T (269 regimens; 27.5%). Although the NRTI 3TC has been used consistently throughout the time period of interest (108; 86.4% of those starting HAART in 1998 had an initial regimen containing 3TC; in 2004 this was 116; 75.3%), for most other antiretrovirals there have been changes in the prescribing patterns. In 1998, the most common NRTIs after 3TC were d4T and AZT received by 73 (58.4%) and 52 (41.6%) patients

respectively. However, in the 2003 BHIVA guidelines d4T was no longer recommended for inclusion in first line HAART regimens due to a high risk of adverse events including peripheral neuropathy [328;355]. In 2004, after 3TC, the most popular NRTIs were TDF (68 patients [44.2%]) and AZT (116 patients [75.3%]).

Figure 4.3 – NRTIs included in initial antiretroviral regimen amongst those starting HAART from 1998 to 2004



The PIs and NNRTIs included in first line regimens were varied, with ritonavir (both as the main PI and as a boosting agent 294; 30.1%), LPV (220; 22.5%) and NFV (88; 9.0%) being the most common PIs, and EFV (384; 39.3%) the most frequently used NNRTI (Figure 4.4). Overall, the use of the NNRTIs EFV and NVP has increased over time, agreeing with the finding of the increased use of 1NNRTI+2NRTI regimens described in Table 4.4. NVP was included in the first-line regimens of 23.2% of individuals in 1998, increasing to 29.8% by 1999. However, this percentage has since fallen, and in 2004 only 5.2% of individuals included NVP in their first-line regimens. In contrast EFV was barely used in 1998. However, in 1999 22.6% of regimens included this antiretroviral. This proportion has now increased and has remained at approximately 50% since 2000. The most common PIs used in first line HAART regimens were IDV and NFV in 1998; in 2004 this had changed to RTV (often used as a boosting agent), and LPV.

Figure 4.4 – PIs and NNRTIs included in initial antiretroviral regimen from 1998 to 2004

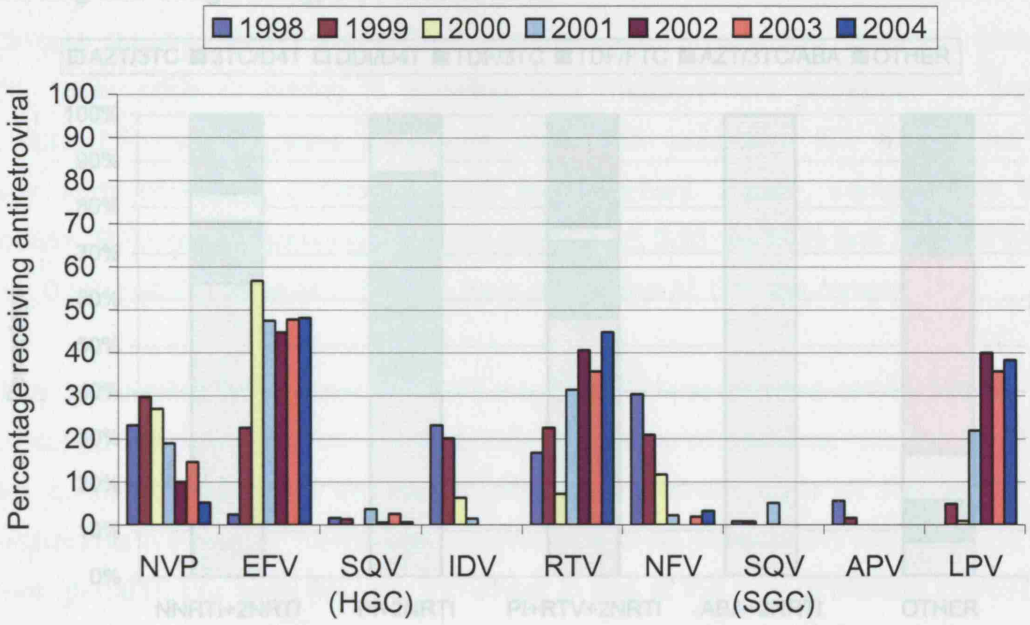
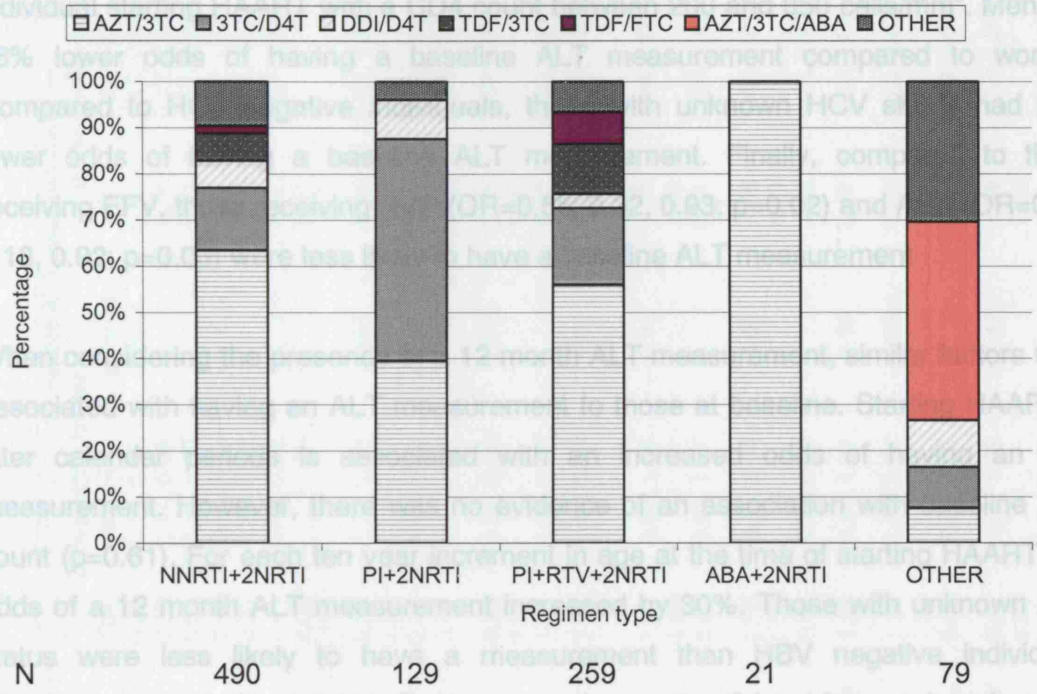


Figure 4.5 shows the most common NRTI backbones used in first line HAART regimens according to the type of regimen, most likely reflecting changes over time in regimen type. When considering the 490 patients who started HAART with 1 NNRTI + 2 NRTI, the most common NRTIs received were AZT with 3TC in 311 (63.5%) cases, followed by 3TC and d4T, received by 66 (13.5%) patients. In contrast, amongst the 129 individuals starting HAART with 1 PI and 2 NRTI, the most common NRTI combinations were d4T with 3TC (62 cases [48.1%]) and AZT with 3TC (51 patients [39.5%]). Initial regimens containing ABA and 2 NRTIs always included AZT and 3TC, most likely as all three antiretrovirals can be taken as Trizivir in one tablet. There is more variety in the NRTI backbones prescribed as part of a PI with RTV and 2 NRTIs regimen; in over half of cases (145/259; 56.0%) the NRTI backbone was AZT and 3TC; d4T and 3TC was received in 43 (16.6%) cases, and TDF and 3TC was prescribed in 28 (10.8%) of cases.

Figure 4.5 – Nucleoside reverse transcriptase inhibitor combinations prescribed according to the regimen type in 1998 to 2004



4.5 Factors associated with monitoring laboratory markers amongst those starting HAART in the Royal Free Cohort

I shall now investigate the factors associated with having a laboratory marker measured in the six-month period before starting HAART (baseline), and ten to fourteen months after starting HAART (12 month measurement). These analyses are not looking at the result of the laboratory marker measurement, but the likelihood of having a measurement compared to no measurement.

Table 4.5 shows the results of logistic regression univariable and multivariable analyses investigating the factors associated with the presence of an ALT measurement at baseline and 12 months after starting HAART amongst previously antiretroviral naïve patients. At baseline, patients were more likely to have an ALT measurement taken if they started HAART in later calendar years (p=0.06). For example, after adjusting for other factors, an individual starting HAART in 1998 had 61% lower odds of having a baseline ALT measurement than someone starting HAART in 2004 (adjusted OR=0.39; 95% CI 0.17, 0.89). An individual starting HAART in 2002 had 55% lower odds of having a baseline ALT measurement than someone starting HAART in 2004. The baseline CD4 count was also associated with the presence of a baseline ALT measurement (p=0.009). After adjusting for potential confounders, those with a baseline CD4 count of <200 cells/mm^3 had a 45% reduction in the odds of

having a baseline ALT measurement and those with a baseline CD4 count ≥500 cells/mm^3 had 41% reduction in the odds of having a measurement compared to an individual starting HAART with a CD4 count between 200 and 350 cells/mm^3. Men had 66% lower odds of having a baseline ALT measurement compared to women. Compared to HCV negative individuals, those with unknown HCV status had 90% lower odds of having a baseline ALT measurement. Finally, compared to those receiving EFV, those receiving NVP (OR=0.55; 0.32, 0.93; p=0.02) and ABA (OR=0.34; 0.13, 0.92; p=0.03) were less likely to have a baseline ALT measurement

When considering the presence of a 12 month ALT measurement, similar factors were associated with having an ALT measurement to those at baseline. Starting HAART in later calendar periods is associated with an increased odds of having an ALT measurement. However, there was no evidence of an association with baseline CD4 count (p=0.61). For each ten year increment in age at the time of starting HAART, the odds of a 12 month ALT measurement increased by 30%. Those with unknown HBV status were less likely to have a measurement than HBV negative individuals (OR=0.53; 0.32, 0.88; p=0.03). Furthermore, those who did not have a baseline ALT measurement had 74% lower odds than those with a normal baseline ALT measurement <40 IU/L. Those who had a high baseline ALT level >40 IU/L had similar odds of having a 12 months ALT measurement, (OR=0.88; 0.58, 1.34) compared to those with lower baseline ALT levels.

As with ALT measurements, later calendar years are associated with an increased chance of having a total cholesterol measurement at baseline (p<0.0001; Table 4.6). For example, someone starting HAART in 1998 had 95% reduction in odds of having a baseline total cholesterol measurement than someone starting HAART in 2004 (OR=0.05; 95% CI 0.02, 0.10), and someone starting HAART in 2002 had 79% reduction in odds. Older age was associated with a higher chance of having a total cholesterol measurement (OR=1.30 per 10 years older; 1.09, 1.56; p=0.004). Compared to HCV negative patients, those with unknown HCV status were less likely to have a measurement (OR=0.23; 0.12, 0.43). The PI/NNRTI included in the HAART regimen was associated with the presence of a baseline total cholesterol measurement in a multivariable analysis. In particular, compared to those receiving a regimen where the only PI or NNRTI used was EFV, those on regimens containing LPV (1.65; 1.07, 2.55) or other regimens (1.94; 1.15, 3.27) were more likely to have a baseline total cholesterol measurement.

Table 4.5 – Results from univariable and multivariable logistic regression analysis investigating factors associated with presence of ALT measurements at the time of starting HAART and 12 months after starting HAART

		Baseline ALT measurement						12 month ALT measurement					
		Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
Variable		**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**P**
Calendar year	*1998*	0.32	0.17, 0.62	0.009	0.39	0.17, 0.89	0.11	0.25	0.14, 0.45	0.0001	0.31	0.15, 0.66	0.02
	1999	0.31	0.16, 0.59		0.33	0.16, 0.69		0.26	0.14, 0.48		0.34	0.17, 0.68	
	2000	0.44	0.22, 0.89		0.52	0.24, 1.11		0.32	0.18, 0.60		0.42	0.21, 0.83	
	2001	0.35	0.18, 0.67		0.40	0.20, 0.80		0.45	0.25, 0.83		0.61	0.31, 1.18	
	2002	0.39	0.21, 0.73		0.45	0.23, 0.88		0.35	0.20, 0.62		0.39	0.21, 0.73	
	2003	0.53	0.28, 1.03		0.56	0.28, 1.12		0.41	0.23, 0.74		0.47	0.25, 0.87	
	2004	1.00	-		1.00	-		1.00	-		1.00	-	
Pre-HAART CD4 count (cells/mm^3)	*<200*	0.62	0.43, 0.89	0.01	0.55	0.36, 0.82	0.009	0.97	0.70, 1.34	0.56	1.10	0.76, 1.59	0.61
	200-350	1.00	-		1.00	-		1.00	-		1.00	-	
	350-500	1.15	0.60, 2.20		1.18	0.59, 2.37		0.80	0.49, 1.33		0.78	0.45, 1.36	
	500+	0.53	0.29, 0.97		0.59	0.30, 1.15		0.71	0.41, 1.24		0.84	0.44, 1.59	
Viral load (log copies/ml)	*<4*	1.13	0.62, 2.08	0.82	1.19	0.59, 2.39	0.78	0.59	0.36, 0.99	0.006	0.66	0.37, 1.18	0.05
	4-5	1.00	-		1.00	-		1.00	-		1.00	-	
	5+	1.10	0.79, 1.54		1.13	0.77, 1.65		1.27	0.94, 1.72		1.30	0.92, 1.84	
Pre-HAART age	*10 years older*	0.99	0.83, 1.18	0.91	1.18	0.96, 1.45	0.11	1.24	1.05, 1.47	0.01	1.30	1.06, 1.58	0.01
Ethnicity	*White*	0.85	0.59, 1.21	0.12	1.02	0.59, 1.77	0.21	1.21	0.88, 1.66	0.23	1.17	0.81, 1.93	0.60
	Black African	1.00	-		1.00	-		1.00	-		1.00	-	
	Other	0.61	0.37, 0.98		0.67	0.38, 1.17		0.88	0.57, 1.37		0.91	0.54, 1.54	
Risk group	*Homosexual*	0.89	0.64, 1.22	0.29	1.15	0.66, 2.01	0.39	1.26	0.94, 1.68	0.01	1.06	0.63, 1.81	0.22
	Heterosexual	1.00	-		1.00	-		1.00	-		1.00	-	
	Other	0.58	0.29, 1.16		0.64	0.27, 1.54		0.51	0.27, 0.95		0.54	0.24, 1.22	
Gender	*Male vs female*	0.52	0.35, 0.77	0.001	0.34	0.20, 0.58	<0.0001	1.15	0.84, 1.57	0.38	0.97	0.60, 1.56	0.89
HBV status	*Positive*	1.52	0.67, 3.43	0.0001	1.68	0.72, 3.96	0.06	1.35	0.66, 2.74	<0.0001	1.26	0.59, 2.67	0.03
	Negative	1.00	-		1.00	-		1.00	-		1.00	-	
	Unknown	0.49	0.35, 0.70		2.17	1.08, 4.39		0.38	0.27, 0.52		0.53	0.32, 0.88	
HCV status	*Positive*	1.39	0.72, 2.69	<0.0001	1.78	0.85, 3.75	<0.0001	0.83	0.50, 1.38	<0.0001	1.04	0.56, 1.91	0.79
	Negative	1.00	-		1.00	-		1.00	-		1.00	-	
	Unknown	0.23	0.15, 0.34		0.10	0.05, 0.22		0.33	0.22, 0.49		0.81	0.44, 1.51	

		Baseline ALT measurement						12 month ALT measurement					
		Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
Variable		**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**P**
PI/NNRTI received	*NVP only*	0.58	0.37, 0.90	0.01	0.55	0.32, 0.93	0.02	0.62	0.41, 0.93	0.02	1.08	0.66, 1.75	0.78
	IDV only	0.66	0.28, 1.53	0.33	0.60	0.21, 1.73	0.35	1.01	0.44, 2.33	0.99	2.16	0.76, 6.17	0.15
	EFV only	1.00	-		1.00	-		1.00	-		1.00	-	
	NFV only	0.91	0.50, 1.68	0.77	0.90	0.43, 1.90	0.78	0.55	0.33, 0.92	0.02	0.84	0.44, 1.59	0.59
	LPV/r only	1.09	0.70, 1.70	0.70	0.99	0.60, 1.62	0.95	1.43	0.94, 2.16	0.09	1.42	0.89, 2.27	0.14
	ABA only	0.41	0.17, 0.96	0.04	0.34	0.13, 0.92	0.03	0.45	0.20, 1.02	0.05	1.00	0.39, 2.52	0.99
	Other	0.75	0.45, 1.25	0.27	0.77	0.43, 1.40	0.40	0.88	0.55, 1.41	0.60	1.19	0.68, 2.09	0.54
Baseline ALT (IU/L)	*Missing*	-	-	-	-	-	-	0.24	0.17, 0.33	<0.0001	0.26	0.17, 0.38	<0.0001
	<40	1.00	-		1.00	-		1.00	-		1.00	-	
	High >40	-	-	-	-	-	-	0.98	0.67, 1.43		0.88	0.58, 1.34	

OR=odds ratio; 95% CI=95% confidence interval; HBV=hepatitis B virus; HCV=hepatitis C virus; PI=protease inhibitor; NNRTI=non nucleoside reverse transcriptase inhibitor

After 12 months of HAART, several factors were associated with having a total cholesterol measurement in a multivariable analysis (Table 4.6). Again, those starting HAART at later time periods were more likely to have a measurement. For example, an individual starting HAART in 1998 had an 82% reduction in odds of having a 12 month total cholesterol measurement compared to someone starting HAART in 2004. For each ten year increment in age, the odds of having a 12 month total cholesterol measurement increased by 24%. Compared to someone with a normal total cholesterol level <6.2 mmol/l at baseline, someone with a missing baseline value had a 63% reduction in the odds of a 12 month measurement (0.37; 0.27, 0.50), and someone with a high baseline measurement had a 70% reduction in their odds (0.30; 0.11, 0.81; p<0.0001).

Table 4.7 shows factors associated with the presence of a test for glucose at baseline and after 12 months of HAART. In a multivariable analysis, calendar year was associated with the presence of a baseline glucose level. For example, someone starting HAART in 1998 had a 93% reduction in their odds of having a baseline glucose measurement compared to someone starting HAART in 2004. Independently, each ten year increment in age at the time of starting HAART was associated with an 18% increase in the odds of a measurement (95% CI 0.99, 1.41; p=0.06). Compared to patients known to be HCV negative, those with unknown HCV status were less likely to have a measurement than HCV negative individuals (OR=0.30; 0.16, 0.55). Compared to a regimen containing EFV, those receiving LPV were more likely to have a baseline total glucose measurement (OR=1.63; 95% CI 1.06, 2.50).

After 12 months of HAART, several factors were found to be associated with the presence of a glucose measurement. Later calendar years were associated with greater monitoring. For example an individual starting HAART in 2002 had 56% lower odds of having a 12 month glucose measurement compared to an individual starting HAART in 2004. For each ten year increment in age, the odds of having a 12 month measurement increased by 21%. Compared to those with a low/normal glucose measurement <4.5 mmol/l at baseline, those with a missing baseline glucose measurement had around half the odds of having a 12 month measurement (0.48; 0.33, 0.68).

Table 4.6 – Results from univariable and multivariable logistic regression analysis investigating factors associated with presence of total cholesterol measurements at the time of starting HAART and 12 months after starting HAART

		Baseline total cholesterol measurement						12 month total cholesterol measurement					
		Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
Variable		**OR**	**95% CI**	**p**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**P**
Calendar year	*1998*	0.06	0.03, 0.11	<0.0001	0.05	0.02, 0.10	<0.0001	0.11	0.06, 0.20	<0.0001	0.18	0.08, 0.38	<0.0001
	1999	0.11	0.06, 0.20		0.10	0.05, 0.19		0.10	0.06, 0.18		0.15	0.08, 0.30	
	2000	0.17	0.09, 0.30		0.17	0.09, 0.32		0.12	0.07, 0.22		0.17	0.09, 0.34	
	2001	0.14	0.08, 0.25		0.14	0.08, 0.26		0.23	0.13, 0.41		0.36	0.19, 0.67	
	2002	0.21	0.12, 0.36		0.21	0.11, 0.38		0.33	0.19, 0.58		0.46	0.25, 0.85	
	2003	0.60	0.32, 1.12		0.65	0.34, 1.23		0.36	0.20, 0.65		0.42	0.23, 0.78	
	2004	1.00	-		1.00	-		1.00	-		1.00	-	
Pre-HAART CD4 count (cells/mm^3)	*<200*	0.87	0.65, 1.17	0.42	0.89	0.63, 1.26	0.27	0.89	0.66, 1.20	0.74	0.89	0.63, 1.26	0.59
	200-350	1.00	-		1.00	-		1.00	-		1.00	-	
	350-500	1.24	0.77, 2.01		1.52	0.87, 2.67		0.78	0.49, 1.24		0.69	0.40, 1.20	
	500+	1.07	0.63, 1.81		1.15	0.62, 2.14		0.90	0.53, 1.52		0.78	0.42, 1.47	
Pre-HAART Viral load (log copies/ml)	*<4*	0.98	0.59, 1.62	0.67	0.96	0.53, 1.73	0.78	0.67	0.41, 1.10	0.21	0.83	0.46, 1.49	0.36
	4-5	1.00	-		1.00	-		1.00	-		1.00	-	
	5+	0.88	0.67, 1.17		1.11	0.80, 1.55		1.02	0.77, 1.35		1.20	0.85, 1.68	
Pre-HAART Age	*10 years older*	1.22	1.05, 1.42	0.009	1.30	1.09, 1.56	0.004	1.24	1.06, 1.44	0.007	1.24	1.03, 1.49	0.02
Ethnicity	*White*	0.92	0.69, 1.23	0.67	0.83	0.51, 1.35	0.59	1.15	0.86, 1.53	0.22	0.95	0.58, 1.54	0.35
	Black African	1.00	-		1.00	-		1.00	-		1.00	-	
	Other	0.83	0.55, 1.26		0.77	0.47, 1.28		0.83	0.55, 1.25		0.71	0.43, 1.18	
Risk group	*Homosexual*	1.08	0.83, 1.40	0.68	0.75	0.48, 1.17	0.40	1.37	1.05, 1.78	0.002	0.92	0.59, 1.45	0.18
	Heterosexual	1.00	-		1.00	-		1.00	-		1.00	-	
	Other	0.84	0.45, 1.57		1.38	0.83, 2.28		0.50	0.27, 0.94		1.37	0.82, 2.28	
Gender	*Male vs female*	0.96	0.72, 1.28	0.78	0.75	0.48, 1.17	0.21	1.28	0.96, 1.71	0.09	0.92	0.59, 1.45	0.72
HBV status	*Positive*	1.37	0.75, 2.51	<0.0001	1.26	0.64, 2.50	0.53	1.34	0.72, 2.49	<0.0001	1.26	0.62, 2.54	0.21
	Negative	1.00	-		1.00	-		1.00	-		1.00	-	
	Unknown	0.49	0.35, 0.67		1.28	0.78, 2.09		0.40	0.29, 0.55		0.67	0.41, 1.10	
HCV status	*Positive*	0.77	0.48, 1.22	<0.0001	0.88	0.51, 1.52	<0.0001	0.66	0.42, 1.04	<0.0001	0.84	0.47, 1.48	0.58
	Negative	1.00	-		1.00	-		1.00	-		1.00	-	
	Unknown	0.34	0.23, 0.51		0.23	0.12, 0.43		0.43	0.29, 0.64		0.75	0.40, 1.40	

		Baseline total cholesterol measurement						12 month total cholesterol measurement					
		Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
Variable		**OR**	**95% CI**	**p**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**P**
NNRTI /PI	*NVP only*	0.52	0.36, 0.77	0.0009	0.93	0.59, 1.47	0.75	0.43	0.30, 0.64	<0.0001	0.85	0.54, 1.34	0.48
	EFV only	1.00	-		1.00	-		1.00	-		1.00	-	
	IDV only	0.25	0.11, 0.56	0.0007	1.16	0.44, 3.04	0.77	0.87	0.41, 1.85	0.72	2.88	1.14, 7.31	0.03
	NFV only	0.50	0.31, 0.82	0.006	1.50	0.80, 2.81	0.21	0.33	0.20, 0.55	<0.0001	0.75	0.40, 1.41	0.37
	LPV/r only	1.94	1.32, 2.83	0.0007	1.65	1.07, 2.55	0.02	1.94	1.31, 2.87	0.0009	1.43	0.92, 2.24	0.12
	ABA only	0.66	0.29, 1.50	0.32	0.89	0.36, 2.18	0.79	0.37	0.16, 0.84	0.02	0.51	0.20, 1.29	0.16
	Other	0.90	0.59, 1.38	0.62	1.94	1.15, 3.27	0.01	0.97	0.63, 1.50	0.89	1.67	0.98, 2.86	0.06
Baseline total cholesterol (mmol/l)	*Missing*	-	-	-	-	-	-	0.25	0.20, 0.33	<0.0001	0.37	0.27, 0.50	<0.0001
	<6.2	-	-	-	-	-	-	1.00	-		1.00	-	
	High >6.2	-	-	-	-	-	-	0.34	0.14, 0.83		0.30	0.11, 0.81	

Table 4.7 – Results from univariable and multivariable logistic regression analysis investigating factors associated with presence of glucose measurements at the time of starting HAART and 12 months after starting HAART

		Baseline glucose measurement						12 month glucose measurement					
		Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
Variable		**OR**	**95% CI**	**p**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**P**
Calendar year (2004)	*1998*	0.07	0.04, 0.13	<0.0001	0.07	0.03, 0.15	<0.0001	0.11	0.06, 0.20	<0.0001	0.17	0.08, 0.36	<0.0001
	1999	0.12	0.06, 0.21		0.11	0.06, 0.22		0.10	0.05, 0.18		0.14	0.07, 0.27	
	2000	0.23	0.13, 0.43		0.26	0.13, 0.50		0.12	0.07, 0.22		0.16	0.08, 0.30	
	2001	0.14	0.08, 0.26		0.16	0.09, 0.30		0.18	0.10, 0.32		0.26	0.14, 0.48	
	2002	0.17	0.10, 0.31		0.19	0.11, 0.35		0.31	0.18, 0.55		0.44	0.24, 0.82	
	2003	0.49	0.27, 0.90		0.54	0.29, 1.01		0.37	0.21, 0.67		0.67	0.39, 1.15	
	2004	1.00	-		1.00	-		1.00	-		1.00	-	
Pre-HAART CD4 count (cells/mm^3)	*<200*	1.04	0.64, 1.68	0.15	0.78	0.56, 1.10	0.11	0.86	0.64, 1.16	0.56	0.83	0.45, 1.54	0.52
	200-350	1.00	-		1.00	-		1.00	-		1.00	-	
	350-500	0.60	0.36, 1.01		1.15	0.66, 1.98		0.72	0.45, 1.15		0.85	0.61, 1.21	
	500+	0.80	0.59, 1.07		0.56	0.31, 1.02		0.86	0.51, 1.45		1.16	0.83, 1.61	
Pre-HAART Viral load (log copies/ml)	*<4*	1.13	0.68, 1.88	0.79	1.29	0.71, 2.32	0.64	0.62	0.38, 1.01	0.13	0.63	0.35, 1.12	0.11
	4-5	1.00	-		1.00	-		1.00	-		1.00	-	
	5+	0.96	0.73, 1.27		1.12	0.81, 1.55		0.98	0.74, 1.30		1.16	0.83, 1.61	
Pre-HAART Age	*10 years older*	1.15	0.99, 1.34	0.07	1.18	0.99, 1.41	0.06	1.20	1.03, 1.39	0.02	1.21	1.01, 1.45	0.04
Ethnicity	*White*	0.85	0.63, 1.13	0.26	0.85	0.53, 1.36	0.27	1.03	0.77, 1.38	0.27	0.93	0.58, 1.49	0.30
	Black African	1.00	-		1.00	-		1.00	-		1.00	-	
	Other	0.72	0.7, 1.09		0.67	0.41, 1.09		0.76	0.50, 1.15		0.70	0.43, 1.14	
Risk group	*Homosexual*	1.03	0.79, 1.34	0.23	0.74	0.48, 1.16	0.42	0.49	0.26, 0.92	0.01	0.75	0.48, 1.17	0.22
	Heterosexual												
	Other	0.60	0.32, 1.13		1.02	0.47, 2.22		1.21	0.93, 1.58		1.34	0.81, 2.21	
Gender	*Male vs female*	0.91	0.68, 1.21	0.50	0.74	0.48, 1.16	0.19	1.06	0.80, 1.42	0.69	0.75	0.48, 1.17	0.21
HBV status	*Positive*	1.67	0.89, 3.14	<0.0001	1.62	0.81, 3.23	0.31	1.43	0.77, 2.66	<0.0001	1.36	0.68, 2.75	0.31
	Negative	1.00	-		1.00	-		1.00	-		1.00	-	
	Unknown	0.54	0.40, 0.75		1.23	0.75, 2.01		0.44	0.32, 0.60		0.75	0.46, 1.22	

		Baseline glucose measurement						12 month glucose measurement					
		Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
Variable		**OR**	**95% CI**	**p**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**P**
HCV status	*Positive*	0.68	0.43, 1.09	<0.0001	0.84	0.49, 1.44	0.0006	0.70	0.44, 1.10	0.0003	0.93	0.53, 1.62	0.49
	Negative	1.00	-		1.00	-		1.00	-		1.00	-	
	Unknown	0.39	0.26, 0.58		0.30	0.16, 0.55		0.46	0.31, 0.68		0.69	0.37, 1.27	
NNRTI /PI	*NVP only*	0.49	0.34, 0.72	0.0003	0.80	0.51, 1.26	0.34	0.45	0.31, 0.66	<0.0001	0.87	0.55, 1.36	0.53
	EFV only	1.00	-		1.00	-		1.00	-		1.00	-	
	IDV only	0.24	0.11, 0.53	0.0004	0.83	0.32, 2.16	0.70	0.65	0.31, 1.35	0.24	1.90	0.76, 4.72	0.17
	NFV only	0.64	0.39, 1.04	0.07	1.54	0.82, 2.88	0.18	0.45	0.27, 0.73	0.001	0.94	0.50, 1.75	0.84
	LPV only	1.65	1.13, 2.40	0.009	1.63	1.06, 2.50	0.03	2.19	1.48, 3.24	<0.0001	1.62	1.05, 2.53	0.03
	ABA only	0.45	0.20, 1.03	0.06	0.61	0.25, 1.48	0.27	0.40	0.18, 0.92	0.03	0.63	0.25, 1.59	0.32
	Other	0.85	0.55, 1.30	0.45	1.62	0.97, 2.72	0.07	0.99	0.64, 1.52	0.97	1.66	0.98, 2.80	0.06
Baseline glucose	*Missing*	-	-	-	-	-	-	0.31	0.23, 0.42	<0.0001	0.48	0.33, 0.68	<0.0001
(mmol/l)	*<4.5 mmol/l*	-	-	-	-	-	-	1.00	-		1.00	-	
	High>4.5 mmo/l	-	-	-	-	-	-	1.11	0.78, 1.60		1.24	0.83, 1.84	

In a multivariable model, calendar year was strongly associated with the presence of a baseline HDL measurement (p<0.0001; Table 4.8). The model particularly showed the odds of having a HDL cholesterol measurement in early years was very small, with someone starting HAART in 1999 having just a twentieth of the odds of having a baseline HDL measurement compared to someone starting HAART in 2004 (OR=0.05; 95% CI 0.03, 0.09). Independently of this, each 10 year increment in age at the time of starting HAART was associated with a 25% increase in the odds of having a baseline HDL cholesterol measurement (OR=1.25; 95% CI 1.02, 1.52; p=0.03).

When considering factors associated with a HDL cholesterol measurement at 12 months, those starting HAART in later calendar periods having a much greater chance of having a measurement than those starting HAART in earlier years. For every additional 10 years of age when starting HAART, the odds of having a 12 months HDL cholesterol measurement increased by 24% (1.24; 1.03, 1.50; p=0.03). Compared to those with an HDL cholesterol measurement >1.0 mmol/l at baseline, those without a baseline measurement were less likely to have a 12 month measurement (OR=0.53; 95% CI 0.34, 0.83).

In summary, this section has shown that there have been dramatic changes over time in the frequency of monitoring of laboratory markers often used as surrogates for antiretroviral-related toxicities. Over the same time period, there have been changes in the antiretrovirals used in first-line HAART regimens, and in the characteristics of patients. Therefore, different demographic groups and those receiving different antiretrovirals are likely to have been subject to different frequency of monitoring patterns. When considering the factors associated with measurements of laboratory markers at baseline and at 12 months, after adjusting for calendar year few factors were associated with presence of a measurement.

Table 4.8 – Results from univariable and multivariable logistic regression analysis investigating factors associated with presence of HDL cholesterol measurements at the time of starting HAART and 12 months after starting HAART

		Baseline HDL cholesterol measurement						12 month HDL cholesterol measurement					
		Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
Variable		**OR**	**95% CI**	**p**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**P**
Calendar year	*1998*	0.01	0.00, 0.01	<0.0001	0.03	0.00, 0.01	<0.0001	0.02	0.01, 0.04	<0.0001	0.03	0.01, 0.08	<0.0001
(2004)	*1999*	0.02	0.01, 0.04		0.01	0.01, 0.03		0.04	0.02, 0.08		0.07	0.03, 0.14	
	2000	0.07	0.04, 0.13		0.07	0.03, 0.13		0.06	0.03, 0.11		0.09	0.04, 0.18	
	2001	0.07	0.04, 0.13		0.07	0.04, 0.12		0.11	0.06, 0.20		0.16	0.08, 0.30	
	2002	0.08	0.05, 0.14		0.08	0.05, 0.15		0.29	0.16, 0.51		0.43	0.23, 0.81	
	2003	0.53	0.29, 0.98		0.60	0.32, 1.14		0.33	0.18, 0.59		0.37	0.20, 0.68	
	2004	1.00	-		1.00	-		1.00	-		1.00	-	
Pre-HAART CD4	*<200*	0.81	0.60, 1.08	0.40	0.95	0.64, 1.41	0.58	0.83	0.62, 1.10	0.32	0.81	0.56, 1.17	0.50
count (cells/mm^3)	*200-350*	1.00	-		1.00	-		1.00	-		1.00	-	
	350-500	1.03	0.65, 1.63		1.39	0.71, 2.69		0.73	0.46, 1.16		0.76	0.42, 1.40	
	500+	0.79	0.47, 1.33		0.80	0.40, 1.56		1.11	0.67, 1.86		1.15	0.59, 2.24	
Pre-HAART Viral	*<4*	0.98	0.60, 1.60	0.03	1.11	0.57, 2.15	0.91	0.68	0.42, 1.12	0.17	0.71	0.38, 1.32	0.28
load (log copies/ml)	*4-5*	1.00	-		1.00	-		1.00	-		1.00	-	
	5+	0.68	0.52, 0.90		0.96	0.66, 1.40		0.80	0.61, 1.05		1.17	0.81, 1.67	
Pre-HAART Age	*10 years older*	1.21	1.04, 1.40	0.01	1.25	1.02, 1.52	0.03	1.23	1.06, 1.43	0.006	1.24	1.03, 1.50	0.03
Ethnicity	*White*	0.87	0.65, 1.16	0.41	0.87	0.49, 1.52	0.23	0.89	0.59, 1.34	0.80	1.05	0.62, 1.77	0.61
	Black African	1.00	-		1.00	-		1.00	-		1.00	-	
	Other	0.77	0.50, 1.16		0.61	0.34, 1.10		1.01	0.76, 1.34		0.82	0.48, 1.39	
Risk group	*Homosexual*	1.10	0.84, 1.42	0.15	1.56	0.86, 2.79	0.34	1.13	0.87, 1.46	0.0001	0.82	0.51, 1.33	0.005
	Heterosexual	1.00	-		1.00	-		1.00	-		1.00	-	
	Other	0.57	0.29, 1.12		1.34	0.51, 3.48		0.19	0.08, 0.44		1.15	0.66. 2.00	
Gender	*Male vs female*	1.03	0.78, 1.38	0.83	0.81	0.48, 1.35	0.41	1.08	0.81, 1.44	0.59	0.82	0.51, 1.33	0.42
HBV status	*Positive*	1.57	0.89, 2.76	<0.0001	1.51	0.72, 3.15	0.04	0.94	0.54, 1.65	<0.0001	0.66	0.32, 1.33	0.42
	Negative	1.00	-		1.00	-		1.00	-		1.00	-	
	Unknown	0.34	0.24, 0.49		0.47	0.24, 0.92		0.40	0.28, 0.55		1.16	0.67, 2.02	
HCV status	*Positive*	0.62	0.38, 1.00	0.0005	0.83	0.42, 1.64	0.49	0.51	0.32, 0.82	<0.0001	0.80	0.42, 1.52	0.01
	Negative	1.00	-		1.00	-		1.00	-		1.00	-	
	Unknown	0.46	0.30, 0.70		0.63	0.28, 1.41		0.42	0.28, 0.63		0.35	0.18, 0.70	

		Baseline HDL cholesterol measurement						12 month HDL cholesterol measurement					
		Univariable analysis			Multivariable analysis			Univariable analysis			Multivariable analysis		
Variable		**OR**	**95% CI**	**p**	**OR**	**95% CI**	**P**	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**P**
NNRTI /PI	*NVP only*	0.42	0.28, 0.63	<0.0001	1.11	0.65, 1.90	0.72	0.30	0.20, 0.46	<0.0001	0.64	0.39, 1.05	0.07
	EFV only	1.00	-		1.00	-		1.00	-		1.00	-	
	*IDV only**	-	-	-	-	-	-	0.22	0.09, 0.52	0.0005	2.21	0.73, 6.67	0.16
	NFV only	0.23	0.13, 0.42	<0.0001	1.78	0.76, 4.17	0.18	0.16	0.09, 0.29	<0.0001	0.65	0.29, 1.42	0.28
	LPV/r only	1.51	1.07, 2.13	0.02	1.15	0.74, 1.79	0.53	2.07	1.43, 3.00	0.0001	1.32	0.85, 2.04	0.21
	ABA only	0.32	0.12, 0.81	0.02	0.49	0.16, 1.47	0.20	0.35	0.15, 0.83	0.02	0.57	0.21, 1.54	0.27
	Other	0.68	0.45, 1.04	0.08	3.02	1.65, 5.52	0.0003	0.56	0.37, 0.86	0.008	1.35	0.77, 2.36	0.29
Baseline HDL (mmol/l)	*Missing*	-	-	-	-	-	-	0.19	0.14, 0.27	<0.0001	0.53	0.34, 0.83	0.0001
	>1.0	-	-	-	-	-	-	1.00	-		1.00	-	
	Low<1.0l	-	-	-	-	-	-	1.14	0.75, 1.76		1.31	0.80, 2.15	

* not included as too few events occurred in this group

4.6 Data simulations investigating the most appropriate toxicity endpoint

The previous sections have shown that differential frequency of monitoring of surrogate laboratory markers of antiretroviral-related toxicities exists in the Royal Free cohort. Calendar time is particularly associated with this, with increased monitoring in more recent years. It is hoped that this increased monitoring will lead to a toxicity being ascertained in a more timely fashion. Although this is beneficial to patients, it does mean that the rates of toxicity are likely to be underestimated in those starting HAART in earlier calendar years, when diagnosis may have been delayed. Therefore, any comparison of the prevalence of toxicities across different calendar time periods is likely to be biased. As I have shown, there have also been changes over time in the specific antiretrovirals prescribed, in the regimen types prescribed, and in the characteristics of patients starting HAART. Thus, any analyses of the potential impact of these factors on the occurrence of antiretroviral-related toxicities are also likely to be affected by this bias.

Therefore, it is important to consider the impact of differential frequency of monitoring when assessing changes in the prevalence of antiretroviral-related toxicities over calendar time. This may also become an issue when comparing the prevalence of toxicities, for example between patients receiving two different antiretrovirals or within two demographic groups, if there are differences in the distribution of calendar time of starting HAART between the two groups. Identifying endpoints for use in analyses that are least affected by these biases would enable us to minimise the impact of differential frequency of monitoring. Therefore, I have used data simulations to investigate the impact, if any, of differential monitoring when investigating whether two different antiretrovirals are associated with a different incidence of antiretroviral-related toxicities, and to identify the analytical methods that are least affected by this bias

4.6.1 Mean time to toxicity event of 26 weeks, mean 500 weeks to toxicity resolution, and no differences between treatment arms

As explained in Section 4.2, I begin by assuming that regimens A and B are associated with the same incidence of toxicity. I have assumed the time to a toxicity occurring after starting HAART is exponentially distributed with a mean of 26 weeks (situation 1 in Table 4.1). This means that the median time to an event occurring is 18 weeks, and by 6 months and one year after starting HAART 63.2% and 86.5% of individuals, respectively, will have experienced an event. Thus, in this situation we are considering a common antiretroviral-related toxicity. I have also assumed that, from the date of a

toxicity occurring, the time to experiencing a resolution of this toxicity is also exponentially distributed, this time with a mean of 500 weeks. Thus, one year after the date of experiencing a toxicity event, 9.9% of individuals will have naturally had a resolution of the toxicity without clinician intervention.

The results of 1000 simulations with the assumptions described above are shown in Table 4.9. Endpoints i and ii consider the association between treatment regimen received and the occurrence of toxicity when no bias caused by frequency of monitoring is present, as the true time of the occurrence of a toxicity event is known. Here, the mean log HR was 0.00, and the associated p-value was <0.05 in 5.3% of situations, confirming that this approach is reasonably unbiased. The minimum log HR obtained in any simulation was –0.22 (HR=0.80), and the maximum value obtained was +0.20 (HR=1.22). When considering the true proportion that were experiencing a toxicity event one year after starting HAART, the mean (range) log OR obtained was +0.01 (-0.54, +0.51), corresponding to an OR of 1.01 (0.58, 1.67); this result was significant at the 5% level in 5.9% of the simulations.

I next assumed that the true time at which the toxicity occurred was no longer known, and instead the toxicity was monitored at each visit, with those receiving regimen B having on average more frequent visits than those receiving regimen A, as described in Subsection 4.2.2. Thus, it was not known whether an individual had experienced a toxicity or not until they had had a measurement taken. The extent to which the estimates obtained from each analytical method were affected by this bias was very different (Table 4.9). The mean log HR/OR obtained for each endpoint, which one would expect to equal zero if the results were unbiased, varied from –0.06 to 0.41. Endpoint 6 (one year measurement meets the definition of a toxicity event; missing=failure) was most affected by the bias of differential frequency of monitoring, with this analysis leading to a p-value <0.05 in 87.0% of simulations, 18 times more frequently than one would expect by chance. Furthermore, this analysis implied that those receiving regimen B had higher odds of experiencing a toxicity event (the log OR was greater than zero) in 100.0% of simulations and the mean log OR was 0.41, (corresponding to an OR of 1.51), with a range of 0.05 to 0.80 (1.05 to 2.23), when we would expect an unbiased estimate to have a mean of zero in this situation. Similar results were obtained when using the same endpoint, but with a missing=excluded approach (endpoint 7). Endpoints 1, 2 and 4 also performed badly, albeit with less extreme results than those seen for endpoints 6 and 7.

Table 4.9 – Results of data simulation investigating impact of frequency of monitoring on results of toxicity analyses: mean of 26 weeks to toxicity event and mean of 500 weeks to resolution: same rate of toxicity for both treatment arms

	Endpoint	Log HR/OR: Regimen B vs. Regimen A*	% runs log HR/OR >0	% runs P<0.05
Not affected by frequency of monitoring				
i	No bias – time to event approach	0.00 -0.22, +0.20	45.9	5.3
ii	No bias – number with an event at 1 year	+0.01 -0.54, +0.51	47.6	5.9
More frequent monitoring amongst those receiving regimen B				
1	Time to experiencing event	+0.15 -0.14, +0.37	48.9	98.1
2	At least 1 measurement in the first year meets definition; M=F	+0.25 -0.31, +0.68	95.1	38.9
3	At least 1 measurement in the first year meets definition; M=E	+0.14 -0.46, +0.70	80.8	15.3
4	1st measurement 6 months-1 year meets definition; M=F	+0.13 -0.38, +0.51	84.2	17.9
5	1st measurement 6 months-1 year meets definition; M=E	-0.06 -0.57, +0.36	34.9	7.7
6	Measurement at 1 year meets definition; M=F	+0.41 0.05, 0.80	100.0	87.0
7	Measurement at 1 year meets definition; M=E	+0.37 -0.10, +0.76	95.5	73.3

M=F missing=failure; M=E missing=excluded; HR hazard ratio; OR odds ratio; %=percentage; *mean (range) values estimated

Although there was still evidence of bias with endpoint 3 (any measurement in the first year of HAART meets the definition of a toxicity event, missing=excluded), there was less bias than that seen for the endpoints previously mentioned. The p-value was <0.05 in 15.3% of simulations, and the odds ratio was >0 for 80.8% of simulated datasets. The mean log OR was 0.14 (corresponding to an OR of 1.14). The least biased results, however, were obtained when using endpoint 5, (the first measurement in the period 6 to 12 months after starting HAART; missing=excluded). Here, in 7.7% of the data simulations the p-value was<0.05 and the mean log odds ratio was –0.06, which is also close to zero as one would expect from an unbiased estimate, although it shows even this method leads to results that are somewhat biased.

4.6.2 Results of data simulations when altering the distribution of time to the occurrence of a toxicity event, and time to the resolution of the event: same rates of event for both treatment arms

I next altered the assumptions in my data simulation model, by changing the distribution of the time to a toxicity event and of the time to resolution of a toxicity event, but still assuming that the rate of toxicity was identical regardless of the regimen received (Table 4.1, situations 1 to 8). The results are summarised in Figures 4.6 and 4.7. The average time to the occurrence of toxicity ranges from a mean of 26 weeks (as in situations 1 and 5) to a mean of 500 weeks (situations 4 and 8). Thus, I am considering a range of scenarios from the situation in which toxicities occur frequently, to the situation in which they are less common. For example, when the mean time to a toxicity occurring is assumed to be 500 weeks, by six months and one year after starting HAART one would expect 5.1% and 9.9% of individuals to have experienced a toxicity event, respectively. I have also considered two values for the mean time to resolution of the toxicity without clinician intervention: 500 weeks (situations 1 to 4) or 200 weeks (situations 5 to 8). In these two situations one would expect 5.1% and 12.2% of individuals to experience a resolution of their toxicity event, respectively, by 6 months after occurrence of the event; at one year these figures are 9.9% and 22.9%.

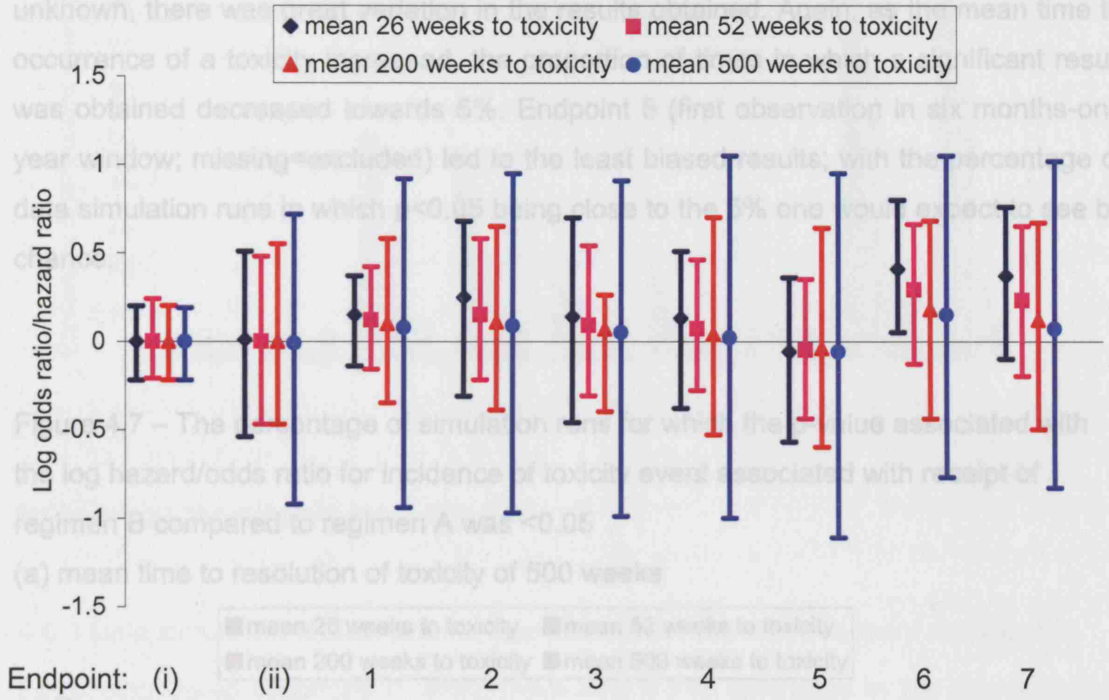
Figure 4.6 shows the mean and range HR/OR obtained when considering different mean times to the occurrence of a toxicity, in the situations in which the mean time to toxicity resolution is 500 weeks (Figure 4.6(a)) and when the mean time to toxicity resolution is 200 weeks (Figure 4.6(b)). As can be seen, when comparing the endpoints in which the true time at which a toxicity occurred is known, (endpoints (i) and (ii)), in all situations the mean log OR/HR was very close to zero, indicating that these methods provide unbiased estimates of the treatment effect. However, the endpoints used when the true time at which a toxicity occurred was not known all lead to biased results. The mean log OR/HRs obtained were positive (with the exception of endpoint 5), suggesting that there is an increased occurrence of toxicity associated with receiving regimen B (the regimen on which more frequent measurements are taken), when in fact the two regimens are equal with respect to toxicity. For all endpoints, the extent to which the estimate of the OR/HR was affected by bias due to frequency of monitoring decreased as the toxicity event became rarer. However, as the toxicity event became rarer, the range of HRs/ORs obtained during the 1000 simulation runs increased. For example, when considering endpoint 6 (measurement at 1 year meets definition; missing=failure), and a mean time to toxicity resolution of 500 weeks, the mean log OR obtained was +0.41 (range +0.02, +0.80) when the mean time to a

toxicity was 26 weeks, but only +0.15 (range –0.07, 1.05) when the mean time to a toxicity was 500 weeks.

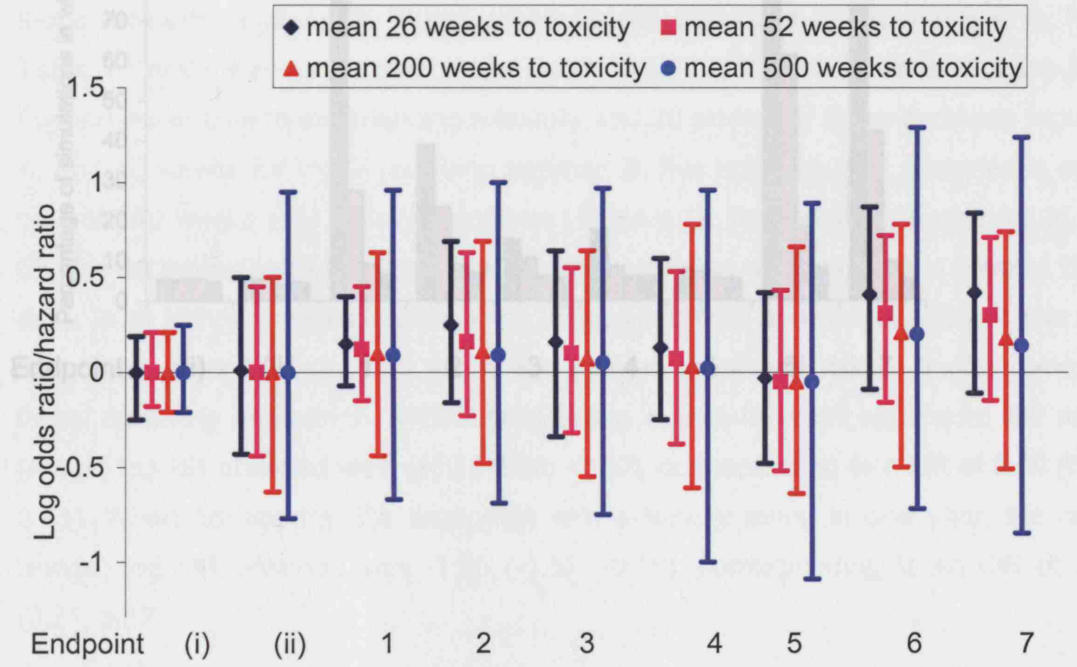
The least biased estimates were obtained when considering endpoint 5 (the first measurement in the first year of HAART meets the definition of a toxicity event, missing=excluded). Using this endpoint, across all situations considered here, the mean log OR obtained varied from –0.06 (OR=0.94) to –0.03 (OR=0.97). Thus, although this endpoint remains biased, the mean parameter estimates obtained are close to zero.

Figure 4.6 – Estimates of mean and range log hazard/odds ratio for incidence of toxicity event associated with receipt of regimen B compared to regimen A
More frequent monitoring amongst those receiving regimen B

(a) mean time to resolution of toxicity of 500 weeks



(b) mean time to resolution of toxicity of 200 weeks

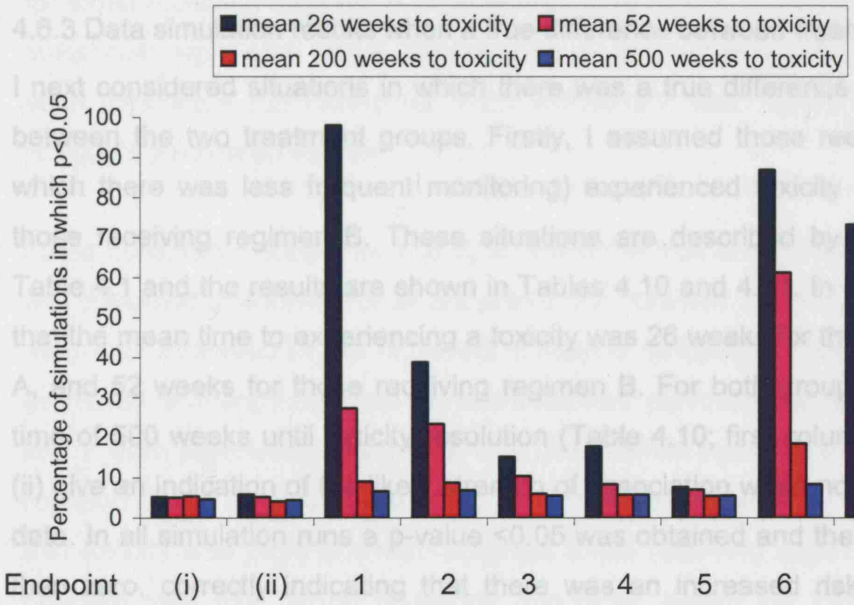


For definitions of endpoints see Table 4.9

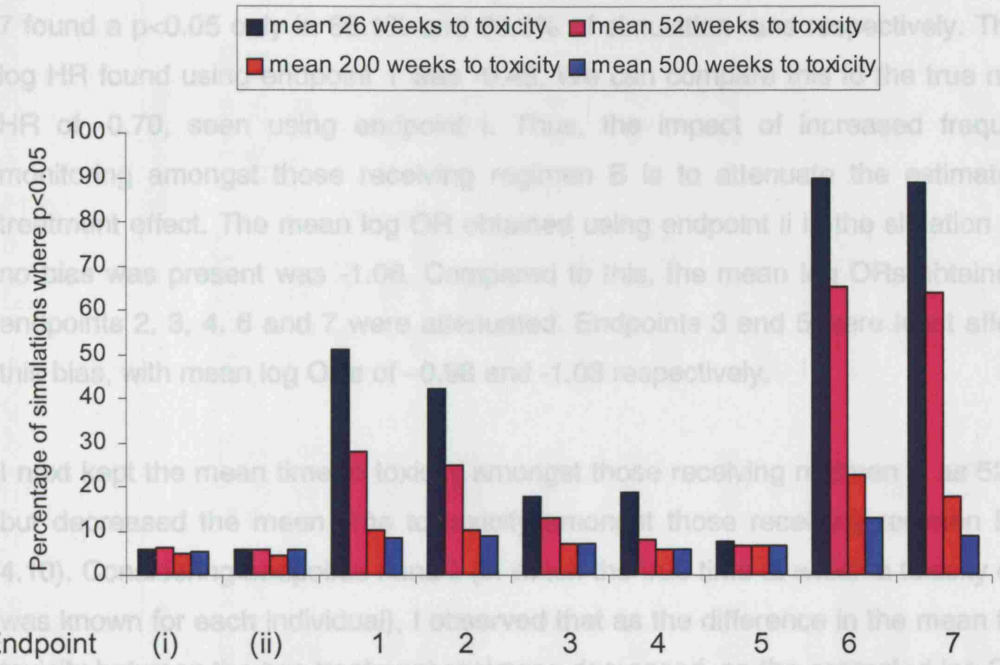
Considering the same situations described in Figure 4.6, Figure 4.7 shows the percentage of simulation runs in which the p-value obtained was <0.05. Again, when considering the endpoints in which the time to toxicity was known, the percentage of simulations in which p<0.05 was very close to 5% in all situations, as one would expect by chance. When considering the endpoints in which the true time of toxicity was unknown, there was great variation in the results obtained. Again, as the mean time to occurrence of a toxicity increased, the proportion of times in which a significant result was obtained decreased towards 5%. Endpoint 5 (first observation in six months-one year window; missing=excluded) led to the least biased results, with the percentage of data simulation runs in which p<0.05 being close to the 5% one would expect to see by chance.

Figure 4.7 – The percentage of simulation runs for which the p-value associated with the log hazard/odds ratio for incidence of toxicity event associated with receipt of regimen B compared to regimen A was <0.05

(a) mean time to resolution of toxicity of 500 weeks



(b) mean time to resolution of toxicity of 200 weeks



for definition of endpoints see Table 4.9

4.6.3 Data simulation results when a true difference between treatment arms exists

I next considered situations in which there was a true difference in the rate of toxicity between the two treatment groups. Firstly, I assumed those receiving regimen A (in which there was less frequent monitoring) experienced toxicity at a faster rate than those receiving regimen B. These situations are described by situations 9 to 14 in Table 4.1 and the results are shown in Tables 4.10 and 4.11. In situation 9, I assumed that the mean time to experiencing a toxicity was 26 weeks for those receiving regimen A, and 52 weeks for those receiving regimen B. For both groups I assumed a mean time of 500 weeks until toxicity resolution (Table 4.10; first column). Endpoints (i) and (ii) give an indication of the likely strength of association when no bias is present in the data. In all simulation runs a p-value <0.05 was obtained and the log OR/HR was less than zero, correctly indicating that there was an increased risk of toxicity amongst those receiving regimen A. When considering a time-to-event approach, the mean (range) log HR obtained was -0.70 (-0.90, -0.50), corresponding to a HR of 0.50 (0.41, 0.61). When considering the proportion with a toxicity event at one year, the mean (range) log OR obtained was -1.06 (-1.57, -0.56), corresponding to an OR of 0.35 (0.21, 0.57).

When the true date of toxicity occurring was not known, the mean log OR/HR obtained using all endpoints was less than zero, correctly identifying a higher rate of toxicity amongst those receiving regimen A. Furthermore, the corresponding p-value was

<0.05 in 100% of data simulations using endpoints 1 to 5. In contrast, endpoints 6 and 7 found a p<0.05 only in 68.1% and 86.8% of simulation runs respectively. The mean log HR found using endpoint 1 was -0.48. We can compare this to the true mean log HR of -0.70, seen using endpoint i. Thus, the impact of increased frequency of monitoring amongst those receiving regimen B is to attenuate the estimate of the treatment effect. The mean log OR obtained using endpoint ii in the situation in which no bias was present was -1.06. Compared to this, the mean log ORs obtained using endpoints 2, 3, 4, 6 and 7 were attenuated. Endpoints 3 and 5 were least affected by this bias, with mean log ORs of –0.98 and -1.03 respectively.

I next kept the mean time to toxicity amongst those receiving regimen A as 52 weeks, but decreased the mean time to toxicity amongst those receiving regimen B (Table 4.10). Considering endpoints i and ii (in which the true time at which a toxicity occurred was known for each individual), I observed that as the difference in the mean time to a toxicity between the two treatment regimens decreased, so the estimated log ORs/HRs tended towards zero (Table 4.10). The percentage of simulation runs for which a p-value of <0.05 was obtained also decreased: for endpoint ii and a mean time to toxicity for those receiving regimen B of 30 weeks, only 30.0% of simulation runs resulted in a statistically significant result. This may be as a result of a lack of power, even though each simulation run contained 1000 individuals, and highlights the importance of adequately powered studies.

I next considered endpoints 1 to 7, in which the true time at which a toxicity occurred was assumed unknown. For all endpoints, the obtained mean log HR/OR was closer to zero than observed in the unbiased situations (endpoints i and ii), and therefore the estimated difference between the two treatment regimens was attenuated. Endpoints 6 and 7 were particularly affected by this bias – when mean time to toxicity for those on regimen B was 30 weeks, the mean log ORs obtained were +0.28 and +0.22, implying that toxicity was more common amongst those receiving regimen B when in fact the opposite was true. In all situations, the least biased log OR was obtained using endpoint 5.

I next considered the situations in which a less frequent toxicity was being considered (Table 4.11). Here, the mean time to a toxicity amongst those receiving regimen A was 200 weeks, and the mean time to toxicity amongst those on regimen B varied from 500 weeks to 250 weeks. The results obtained were comparable to those in Table 4.10. Again, endpoint 5 was least biased by differences in frequency of monitoring, and endpoints 6 and 7 were most strongly affected.

Table 4.10 – Results of data simulations investigating impact of frequency of monitoring on results of toxicity analyses when mean time to toxicity for those receiving regimen A is 26 weeks (mean of 500 weeks to toxicity resolution for both groups)

More frequent monitoring amongst those receiving regimen B

Mean time to toxicity for those receiving regimen B		52 weeks			39 weeks			30 weeks		
	Endpoint	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**
i	No bias– time to event approach	-0.70 -0.90, -0.50	0.0	100.0	-0.41 -0.63, -0.19	0.0	100.0	-0.15 -0.36, +0.09	0.9	57.2
ii	No bias– event at 1 year	-1.06 -1.57, -0.56	0.0	100.0	-0.63 -1.09, -0.18	0.0	99.5	-0.23 -0.73, +0.29	7.1	30.0
1	Time to experiencing event	-0.48 -0.73, -0.21	0.0	100.0	-0.22 -0.43, +0.03	99.8	66.8	+0.02 -0.23, +0.26	38.5	4.7
2	At least 1 measurement in 1st year meets definition; M=F	-0.77 -1.18, -0.35	0.0	100.0	-0.37 -0.76, +0.03	0.2	78.6	+0.02 -0.47, +0.50	55.0	4.6
3	At least 1 measurement in 1st year meets definition; M=E	-0.98 -1.45, -0.55	0.0	100.0	-0.55 -0.97, -0.15	0.0	97.0	-0.12 -0.52, +0.40	78.0	10.9
4	1st measurement 6 months-1 year meets definition; M=F	-0.75 -1.22, -0.31	0.0	100.0	-0.40 -0.84, -0.03	0.0	89.0	-0.06 -0.51, +0.39	32.6	7.4
5	1st measurement 6 months-1 year meets definition; M=E	-1.03 -1.44, -0.52	0.0	100.0	-0.65 -1.06, -0.26	0.0	99.7	-0.27 -0.78, +0.21	2.7	48.4
6	Measurement at 1 year meets definition; M=F	-0.31 -0.85, 0.13	0.9	68.1	+0.01 -0.40, +0.51	45.7	5.4	+0.28 -0.09, +0.68	98.5	58.7
7	Measurement at 1 year meets definition M=E	-0.42 -1.01, -0.01	0.0	86.8	-0.07 -0.08, +0.42	31.4	7.8	+0.22 -0.13, +0.64	94.0	35.6

M=F missing=failure; M=E missing=excluded; %=percentage; * regimen B compared to regimen A: mean (range) values obtained.

Table 4.11 – Results of data simulations investigating impact of frequency of monitoring on results of toxicity analyses when mean time to toxicity for those receiving regimen A is 200 weeks (mean of 500 weeks to toxicity resolution for both groups)

More frequent monitoring amongst those receiving regimen B

	Mean time to toxicity for those receiving regimen B	**500 weeks**			**300 weeks**			**250 weeks**		
	Endpoint	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**
i	No bias– time to event approach	-0.70 -0.96, -0.48	0.0	100.0	-0.41 -0.61, -0.19	0.0	100.0	-0.22 -0.41, -0.02	0	56.9
ii	No bias– event at 1 year	-0.76 -1.26, -0.17	0.0	99.5	-0.45 -0.91, +0.10	0.4	77.1	-0.25 -0.70, +0.23	6.4	36.0
1	Time to experiencing event	-0.59 -1.11, +0.05	0.1	62.8	-0.30 -0.86, +0.26	4.0	45.3	-0.13 -0.55, +0.30	78.8	12.1
2	At least 1 measurement in 1st year meets definition; M=F	-0.64 -1.20, +0.06	0.1	90.6	-0.33 -0.92, +0.27	3.8	47.4	-0.14 -0.61, +0.31	20.7	12.9
3	At least 1 measurement in 1st year meets definition; M=E	-0.68 -1.25, +0.03	0.1	93.9	-0.38 -1.00, +0.21	2.4	55.9	-0.18 -0.66, +0.28	15.4	19.2
4	1st measurement 6 months-1 year meets definition; M=F	-0.69 -1.36, +0.08	0.2	89.4	-0.39 -1.13, +0.21	3.0	49.1	-0.20 -0.79, +0.37	15.3	17.2
5	1st measurement 6 months-1 year meets definition; M=E	-0.78 -1.44, -0.01	0.0	94.2	-0.47 -1.23, +0.11	1.5	64.7	-0.28 -0.88, +0.29	7.1	30.1
6	Measurement at 1 year meets definition; M=F	-0.55 -1.22, +0.08	0.9	77.3	-0.24 -0.79, +0.26	9.6	25.9	-0.06 -0.63, +0.52	38.5	5.5
7	Measurement at 1 year meets definition M=E	-0.61 -1.29, +0.01	0.1	85.6	-0.31 -0.86, +0.23	4.9	38.6	-0.12 -0.69, +0.50	24.7	9.3

M=F missing=failure; M=E missing=excluded; %=percentage; * regimen B compared to regimen A: mean (range) values obtained

I next considered the reverse situation, in which those receiving regimen A (in which there was less frequent monitoring) experienced toxicity at a slower rate than those receiving regimen B. These situations are described in situations 15 to 20 in Table 4.1 and the results are shown in Tables 4.12 and 4.13. When considering a common toxicity (Table 4.12), in which the toxicity occurred after a mean of 26 weeks amongst those receiving regimen B, in all situations and using all endpoints the mean HR/OR obtained was greater than zero, correctly identifying that those receiving regimen B had a higher risk of experiencing an event. When comparing the results of the endpoints to those obtained when no bias was present (endpoints i and ii), the magnitude of the effect associated with receiving regimen B compared to receiving regimen A was generally overestimated. The exception was endpoint 5, which slightly underestimated the effect of treatment regimen B on the occurrence of toxicity.

When considering a less common toxicity (Table 4.13), in which the mean time to a toxicity development was 200 weeks amongst those receiving regimen B, similar results were obtained. Again, the true effect of treatment regimen B compared to treatment regimen A on the occurrence of toxicity was overestimated when compared to the situation in which the true time to toxicity occurrence was known. The exceptions were endpoints 4 and endpoint 5, both of which consider the first measurement within the window of 6 months to 1 year after starting HAART. Here, the impact of regimen B on the occurrence of toxicity was underestimated.

Table 4.12 – Results of data simulations investigating impact of frequency of monitoring on results of toxicity analyses when mean time to toxicity for those receiving regimen B is 26 weeks (mean of 500 weeks to toxicity resolution for both groups)

More frequent monitoring amongst those receiving regimen B

	Mean time to toxicity for those receiving regimen A	**52 weeks**			**39 weeks**			**30 weeks**		
	Endpoint	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**
i	No bias– time to event approach	0.69 0.45, 0.89	100.0	100.0	0.41 0.24, 0.66	100.0	100.0	0.14 -0.04, 0.37	99.2	57.8
ii	No bias– event at 1 year	1.05 0.14, 0.64	100.0	100.0	0.64 0.14, 1.19	100.0	99.0	0.22 -0.19, 0.73	90.7	28.1
1	Time to experiencing event	0.75 0.47, 1.01	100.0	100.0	0.50 0.25, 0.75	100.0	100.0	0.27 0.04, 0.55	100.0	56.8
2	At least 1 measurement in 1st year meets definition; M=F	1.17 0.75, 1.59	100.0	100.0	0.81 0.37, 1.31	100.0	100.0	0.44 0.06, 1.02	100.0	88.2
3	At least 1 measurement in 1st year meets definition; M=E	1.22 0.75, 1.67	100.0	100.0	0.82 0.26, 1.36	100.0	99.9	0.38 -0.15, 1.04	99.5	70.5
4	1st measurement 6 months-1 year meets definition; M=F	0.95 0.54, 1.36	100.0	100.0	0.62 0.25, 1.11	100.0	99.9	0.30 -0.15, 0.66	99.4	65.5
5	1st measurement 6 months-1 year meets definition; M=E	0.92 0.45, 1.40	100.0	100.0	0.55 0.13, 1.08	100.0	97.2	0.16 -0.24, 0.59	86.3	20.5
6	Measurement at 1 year meets definition; M=F	0.99 0.55, 1.36	100.0	100.0	0.74 0.30, 1.20	100.0	100.0	0.51 0.15, 0.91	100.0	97.8
7	Measurement at 1 year meets definition M=E	1.02 0.55, 1.39	100.0	100.0	0.71 0.32, 1.26	100.0	100.0	0.48 0.07, 0.92	100.0	93.5

M=F missing=failure; M=E missing=excluded; %=percentage; * regimen B compared to regimen A: mean (range) values obtained

Table 4.13 – Results of data simulations investigating impact of frequency of monitoring on results of toxicity analyses when mean time to toxicity for those receiving regimen B is 200 weeks (mean of 500 weeks to toxicity resolution for both groups)

More frequent monitoring amongst those receiving regimen B

	Mean time to toxicity for those receiving regimen A	**500 weeks**			**300 weeks**			**250 weeks**		
	Endpoint	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**	**Log HR/OR***	**% runs log HR/OR >0**	**% runs P<0.05**
i	No bias– time to event approach	0.69 0.46, 0.90	100.0	100.0	0.40 0.16, 0.66	100.0	100.0	0.22 0.04, 0.43	0.0	57.9
ii	No bias– event at 1 year	0.76 0.22, 1.40	100.0	99.7	0.45 -0.09, 0.99	99.7	75.2	0.24 -0.23, 0.79	95.2	31.2
1	Time to experiencing event	0.79 0.20, 1.39	100.0	100.0	0.50 -0.01, 1.04	99.8	66.0	0.31 -0.23, 0.90	97.9	47.3
2	At least 1 measurement in 1st year meets definition; M=F	0.84 0.22, 1.50	100.0	99.4	0.54 -0.01, 1.12	99.9	83.7	0.33 -0.26, 0.98	79.7	48.4
3	At least 1 measurement in 1st year meets definition; M=E	0.81 0.19, 1.46	100.0	99.1	0.51 -0.06, 1.07	99.8	78.8	0.30 -0.27, 0.96	96.6	39.9
4	1st measurement 6 months-1 year meets definition; M=F	0.77 0.03, 1.52	100.0	94.1	0.47 -0.21, 1.26	98.3	63.4	0.27 -0.32, 0.99	91.5	28.2
5	1st measurement 6 months-1 year meets definition; M=E	0.70 0.00, 1.44	99.9	89.9	0.40 -0.31, 1.20	97.0	47.6	0.19 -0.40, +0.95	82.8	16.4
6	Measurement at 1 year meets definition; M=F	0.91 0.35, 1.60	100.0	99.6	0.61 -0.01, 1.27	99.9	91.3	0.40 -0.15, 0.99	98.9	61.9
7	Measurement at 1 year meets definition M=E	0.83 0.24, 1.51	100.0	99.2	0.54 -0.05, 1.21	99.7	82.8	0.33 -0.23, 0.93	96.6	46.9

M=F missing=failure; M=E missing=excluded; %=percentage; * regimen B compared to regimen A: mean (range) values obtained

4.6.4 Data simulation results when frequency of monitoring is affected by the presence of a toxicity

Finally, I considered the situation in which those who experienced a toxicity experienced more frequent monitoring. The results are shown in Table 4.14. When considering the situation in which the rate of toxicity was the same in both treatment groups (first two columns), I found that no endpoint gave an unbiased estimate of the treatment effect. The least biased estimates were endpoints 4 and 5. When considering endpoint 4, the mean estimated ORs were +0.09 (when the mean time to a toxicity was 26 weeks) and +0.01 (when the mean time to toxicity was 200 weeks), and the associated p-value was <0.05 on 11.3% and 5.2% of occasions respectively. When considering endpoint 5, the the mean estimated ORs were -0.10 (when the mean time to a toxicity was 26 weeks) and -0.07 (when the mean time to toxicity was 200 weeks), and the associated p-value was <0.05 on 8.8% and 7.1% of occasions respectively. Thus, these estimates resulted in a greater amont of bias compared to when the frequency of monitoring was only affected by the treatment regimen received.

The third and fourth columns of Table 4.14 consider the situations in which a toxicity was more common amongst those receiving regimen A. Using all endpoints, the mean estimated OR/HR was less than zero, correctly identifying a higher level of toxicity amongst those receiving regimen A. However, compared to the mean log OR/HR obtained when no bias was present, the mean estimates were closer to zero than one would expect from an unbiased estimator. The exception was again endpoint 5, which appeared to give least biased estimates of the treatment effect. Finally, the fifth and sixth columns consider the situation in which a toxicity was more common amongst those receiving regimen B. Using all endpoints, the mean estimated OR/HR was greater than zero, correctly identifying a higher level of toxicity amongst those receiving regimen B. However, compared to the mean log OR/HR obtained when no bias was present, the mean estimates from all endpoints, except for endpoint 5, were larger than one would expect from an unbiased estimator, and thus the impact of regimen B on the occurrence of toxicity was overestimated. In contrast endpoint 5 appears to underestimate the impact of regimen B on the occurrence of toxicity.

Table 4.14 – Results of data simulations investigating impact of frequency of monitoring on results of toxicity analyses when the frequency of monitoring is associated with presence of a toxicity

Mean time to toxicity # Endpoint	A: 26 B: 26		A: 200 B: 200		A: 26 B: 52		A: 200 B: 400		A: 52 B: 26		A: 400 B: 200	
	Log HR/OR*	% runs P<0.05	Log HR/OR*	% runs P<0.05	Log HR/OR*	% runs P<0.05	Log HR/OR*	% runs P<0.05	Log HR/OR*	% runs P<0.05	Log HR/OR*	% runs P<0.05
i	0.00 -0.18, +0.17	5.0	0.00 -0.23, +0.21	4.7	-0.69 -0.94, -0.49	100.0	-0.69 -0.93, -0.50	100.0	+0.70 +0.49, +0.89	100.0	+0.69 +0.52, +0.93	100.0
ii	-0.01 -0.55, +0.43	4.7	0.00 -0.47, +0.51	4.8	-1.06 -1.55, -0.59	100.0	-0.75 -1.33, -0.29	99.5	+1.05 -0.50, +1.53	100.0	+0.75 -0.24, +1.36	98.9
1	+0.18 -0.08, +0.41	60.6	+0.11 -0.35, +0.66	11.8	-0.46 -0.73, -0.20	100.0	-0.58 -1.27, -0.03	65.1	+0.79 +0.51, +1.00	100.0	+0.81 +0.14, +1.33	100.0
2	+0.28 -0.56, +0.70	47.0	+0.12 -0.39, +0.74	11.2	-0.76 -1.18, -0.33	100.0	-0.63 -1.34, -0.03	92.5	+1.25 +0.82, +1.66	100.0	+0.86 +0.16, +1.41	99.7
3	+18.6 -0.38, +0.67	21.6	+0.08 -0.44, +0.67	6.9	-0.99 -1.48, -0.52	100.0	-0.68 -1.36, -0.09	95.9	+1.31 +0.86, +1.83	100.0	+0.84 +0.11, +1.37	99.5
4	+0.09 -0.41, +0.48	11.3	+0.01 -0.54, +0.64	5.2	-0.81 -1.19, -0.44	100.0	-0.72 -1.49, -0.14	92.5	0.95 +0.54, +1.34	100.0	+0.74 +0.03, +1.57	93.0
5	-0.10 -0.56, +0.30	8.8	-0.07 -0.65, +0.57	7.1	-1.05 -1.44, -0.59	100.0	-0.81 -1.58, -0.22	97.0	+0.89 +0.47, +1.30	100.0	+0.67 -0.01, +1.51	86.0
6	+0.51 0.00, +0.92	96.4	+0.20 -0.32, +0.75	22.0	-0.33 -0.75, +0.06	73.5	-0.54 -1.18, 0.00	84.1	+1.17 +0.56, +1.62	100.0	+0.92 +0.36, +1.53	100.0
7	+0.43 -0.12, +0.92	82.6	+0.13 -0.40, +0.66	11.7	-0.47 -0.90, -0.03	93.5	-0.61 -1.24, -0.10	91.6	+1.16 +0.56, +1.72	100.0	+0.84 +0.29, +1.43	99.4

M=F missing=failure; M=E missing=excluded; %=percentage; mean (range) values. In all situations mean time to toxicity resolution=500 weeks
* Mean (range) values – comparing regimen B to regimen A. # Mean time to toxicity given in weeks
For definition of endpoints see Table 4.9;

4.7 Discussion

In this Chapter I have investigated the frequency of monitoring of laboratory markers that are used to define, at least in part, HAART-related toxicities. The increase in the monitoring of toxicity-related laboratory markers, in particular ALT, total cholesterol, glucose and HDL cholesterol, has been dramatic. This reflects increased awareness over the time period of the potential toxicities associated with HIV treatment [73;142;145;146], and also a move away from HIV being thought of as a terminal disease to a chronic disease, in which other co-morbidities are being investigated in more detail [86;89;169]. It may therefore be hard to investigate whether there have been any real changes over time in the prevalence of HAART-related toxicities.

Although the frequency of monitoring of all four markers has increased over time, this increase has been particularly dramatic for total cholesterol and HDL cholesterol. Thus, any biases that result from changes in the frequency of monitoring are likely to be most noticeable when considering surrogate endpoints for cardiovascular disease. Indeed, so few patients were monitored for these toxicities in the early years of HAART that it may be impossible to investigate the occurrence of these toxicities in the early years of HAART.

There have been considerable changes over time in the characteristics of patients starting HAART, reflecting changes in the HIV positive population in the UK at the same time [359]. There have been increases in the proportion of women, those of black African ethnicity and those with a heterosexual risk for HIV infection starting HAART. Therefore, one must be aware that any biases associated with calendar year may also affect analyses comparing these demographic factors. Although multivariable analyses can be adjusted for calendar year, and thus hopefully remove some of the associated biases, this adjustment will not generally account for differential frequency of monitoring. Possible biological reasons for real differences between demographic groups, for example as a result of different genetic profiles which may lead to different drug absorption rates and thus different toxicity rates [360], mean that this may be an important issue to investigate. Thus, it is important to identify methods that are least affected by differential frequency of monitoring.

There was also a marked change over time in the antiretrovirals prescribed. With the exception of the NRTI 3TC, which has been widely used throughout the study period (although even here there has been a move towards the use of FTC in the most recent years), the antiretrovirals used in first-line regimens have changed with the introduction

of new antiretrovirals, and as more is known about the efficacy and toxicity profiles of each drug. There are a number of implications of this. Firstly, observational studies that consider the occurrence of HAART-related toxicities as a whole, regardless of the specific antiretrovirals, may consider very different antiretrovirals if the studies are carried out in different calendar years. There may also be potential biases when trying to compare specific antiretrovirals in observational studies, as patients receiving those antiretrovirals that were introduced in earlier time periods are likely to have been monitored less frequently, and therefore less information for these antiretrovirals is likely to be available. Finally, as changes in the demographic profiles of patients occurred simultaneously, it may be hard to disentangle the effects of specific antiretrovirals, demographics and changes over time [361].

My investigations have also demonstrated that the NRTI backbone combinations that have been used as part of different types of treatment regimens are very different. There is evidence from clinical trials with identical NRTI backbones that there may be true differences between antiretrovirals with regard to the prevalence of drug-related toxicities. For example, the 2NN trial, a randomised comparison amongst previously antiretroviral naïve patients comparing the NNRTIs NVP and EFV, used in combination with d4T and 3TC, found that NVP was associated with larger HDL-cholesterol increases and smaller increases in total cholesterol than EFV [362]. Nonetheless, when considering observational data such those considered in this thesis, we must be aware that any differences seen between PIs and NNRTIs may also be due, in part, to the NRTIs received with the drugs. There is evidence that different NRTI drugs may be associated with different toxicity profiles [272]. Thus, when carrying out analyses it is important to describe in detail the drug combinations that patients in the study were receiving and adjust for this wherever possible.

When considering the factors that were associated with the presence of laboratory markers at baseline and after 12 months of HAART using logistic regression models, we found in all cases that calendar year was a strong predictor of the presence of a measurement. After adjusting for this, few other factors were associated with an available measurement. There was evidence that missing unknown HBV and HCV status was also associated with a lower chance of having a measurement taken. However, there was less evidence that having a positive or negative serostatus was associated with the presence of a measurement. I found that older patients were more likely to have a baseline and 12 month total cholesterol and HDL cholesterol measurement taken, which is perhaps to be expected as older age is associated with an increased risk of cardiovascular disease [363]. Excluding this, it is reassuring that

there was no other evidence of an association between any demographic factors and presence of a measurement after adjusting for calendar year.

Korthuis et al investigated the factors associated with the presence of a total cholesterol measurement amongst those starting a PI in the Veterans Administration Study in the USA [364]. They found that 59% of PI-exposed individuals had a lipid screening within the first six months of treatment. Those with a measurement were more likely to be older, with a history of hyperlipidaemia, CVD, diabetes or hypertension, to be of White or Latino ethnicity, and to have a homosexual risk for HIV transmission. The authors considered a two-year window, and therefore, the impact of calendar year was not investigated. Other than this, few studies have considered this issue.

Although we can account for some of the confounding caused by changes over time using standard methods for adjustment in our statistical models, it is harder to adjust for changes in the frequency of monitoring, as those with more frequent monitoring are more likely to be identified as having experienced an event, and also to have random measurements outside of the normal range. Thus, it is important to consider endpoints for our studies that are least affected by frequency of monitoring and will therefore provide unbiased estimates. In this chapter I wished to consider simple endpoints that could be used in analyses, rather than more complicated analytical techniques. The results of my simulations have demonstrated the great extent to which results may be influenced by the frequency of monitoring.

The results of my data simulations suggest that, in general, those endpoints that took a missing=excluded approach generally performed better than those with a missing=failure approach. This is in contrast to RCTs which mostly use a missing=failure approach for their primary endpoint, as required by the USA Food and Drug Administration (FDA) [365]. In the first 20 situations considered in my data simulations the only factor influencing the frequency of monitoring was the treatment regimen received. Therefore, those receiving regimen A, in which less frequent monitoring occurred, were more likely to have a missing value in the windows considered. As a result a missing=failure approach led to the appearance of a higher occurrence of toxicity amongst those on regimen A. As I assumed there was no association between the occurrence of toxicity and frequency of monitoring, those with a missing value were as likely as any patient with a value recorded to have experienced a toxicity or not. Therefore, although a missing=excluded approach reduces the power of any analyses, it is not likely to lead to biased results. This may be

a reasonable approach to take when considering toxicities that do not result in clinical manifestation. For example, changes in total cholesterol, although they lead to an increased risk of cardiovascular disease, are unlikely to manifest themselves as symptomatic illness in the short term. In contrast, changes in laboratory markers that are associated with short-term symptomatic illness may lead to more frequent monitoring. Thus, in my final set of simulations, I considered the situation in which those who experienced a toxicity had more frequent monitoring. As it is difficult to know how much more frequently patients are likely to visit when experiencing a toxicity event, the results of my simulations are dependent upon the assumptions made with regard to the increased frequency of monitoring amongst those who experience a toxicity event. However, under the assumptions made, I found that the missing=excluded approach and missing=failure approach led to comparable results. Thus, in situations in which the frequency of monitoring is likely to be associated with the presence of a HAART-related toxicity, it may be important to consider both approaches and assess whether the results obtained are consistent. In real life situations, the decision to visit the clinic may be more complex than the situation modelled here. Therefore, for each individual situation, we must weigh up the benefits of taking a missing=excluded approach with the benefits of taking a missing=failure approach.

The results from my simulations suggested that taking a fixed time point, such as one year after starting HAART, and investigating the proportion who have experienced a toxicity event led to biased results at this time point. This may be because a relatively narrow window was used (from ten to fourteen months), and therefore those with less frequent monitoring are much less likely to have a measurement recorded within the window of interest. It was particularly concerning that taking the one-year value as the endpoint when considering toxicities was biased even when there was a true difference between treatment arms, to the extent that in certain situations it continued to imply that regimen B was associated with a greater prevalence of toxicity even when those receiving regimen A were at a much higher risk of experiencing a toxicity. The time to a toxicity event was also a biased estimator of treatment effect when there was differential frequency of monitoring. This is perhaps to be expected, as those receiving regimen B were monitored more frequently, and therefore were more likely to be diagnosed with a toxicity at an earlier time point than those receiving regimen A. Use of the proportion with at least one high measurement in the first year of HAART also led to biased results when estimating a treatment effect, although this bias was less marked than that associated with the two approaches previously mentioned. This is likely to be related to the fact that those receiving regimen B had more measurements

recorded, on average, than those on regimen A, and were therefore more likely to have a measurement taken once a toxicity had occurred. The outcome that appeared to be an unbiased estimator of a treatment effect was the first measurement in the period six months to 12 months after starting HAART. Here, when considering situations in which there was no true treatment effect, the log odds ratio was close to zero in nearly all situations, and the p-value was significant at the 5% level in approximately 5% of simulations as one would expect by chance. Once an association between the presence of a toxicity and the frequency of monitoring was assumed, this endpoint was more affected than it had been when this association was not assumed. Nonetheless, of all the endpoints considered it was the one that was least affected by bias and so my simulations would suggest that this is the most appropriate endpoint to use in analyses. The use of a large window, giving individuals an opportunity to have a measurement taken, may lead to less biased results, as those with less frequent measurements have the opportunity to have a measurement recorded. Defining the presence of a toxicity on the basis of the *first* measurement, rather than *any* measurement in the window of interest, means that those individuals with more measurements during the window do not have a greater opportunity to have a measurement taken after a toxicity had occurred.

I wished to consider both common and rare HAART-related toxicity events, as both can occur in real life. I considered rare events by considering a relatively long mean time to occurrence of a toxicity. However, even in the most extreme case for a rare event I considered that 10% of patients experienced a toxicity in the first year of HAART, which perhaps is still quite common. Therefore, further exploratory analyses investigating even more rare events may be an appropriate extension of this work. In my data simulations, I found that as the toxicity became less common, so the endpoints were more able to account for any biases present. Therefore, bias as a result of increased frequency of monitoring may be less important when considering extremely rare events. However, as the most appropriate choice of endpoint does not seem to depend on the true frequency of the toxicity, the first measurement in a large window endpoint appears the most appropriate choice of endpoint.

A closely related issue to frequency of monitoring is that of missing data. In my data simulations, although patients did not necessarily have a measurement recorded during the time period in each of the endpoints being considered, all were assumed to remain under follow-up for the duration of the study. I have also assumed that no factors, other than calendar time (and, by consequence, treatment regimen received) and the presence of a toxicity measurement, were associated with the presence of a

measurement. The results of the logistic regression models in the Royal Free cohort suggested that no other factors were associated with presence of a laboratory marker, although unmeasured factors could be associated. Therefore, it appears reasonable to assume that values could be considered as missing as random, and methods for accounting for data that are not missing at random, such as multiple imputation [366] and sample selection models (used in a different context in Chapter 7 of this thesis) [367], have not been investigated here. Furthermore, I wished to consider simple endpoints to account for differential monitoring. Nonetheless, it is important to consider whether patients are likely to systematically drop out of a study, particularly as a result of experiencing a worse outcome [368], and to consider methods appropriate to account for these if required.

4.8 Summary

In conclusion, this chapter has shown that differential frequency of monitoring is likely to be an important issue when investigating the prevalence and incidence of HAART-related toxicities, mostly attributed to increases in the frequency of monitoring in more recent calendar years. As the demographics of patients starting HAART, and the antiretrovirals used in first line HAART regimens have also changed over time, biased results may be obtained from observational studies if the effects of differential frequency of monitoring are not removed. Choosing an endpoint that is least affected by differences in the frequency of monitoring, such as the first measurement taken in the period 6 to 12 months after starting HAART with a missing=excluded approach, will help to minimise these biases.

Chapter 5 – The impact of endpoint definitions on the observed prevalence of antiretroviral-related toxicity

5.1 Introduction

One of the potential biases identified in Chapter 2 when considering antiretroviral-related toxicities was the difficulty associated with choosing the most appropriate outcome measure for toxicity. Studies have differed in the toxicity definition used, considering different cut-offs, and the use of a single or confirmatory result. The aim of this chapter is to investigate the impact of different definitions on the prevalence of antiretroviral-related toxicity observed. I shall focus here on short-term toxicities that occur in the first year of HAART and on the prevalence (rather than incidence) of toxicity. Firstly, I shall consider the overall prevalence of any toxicity events by considering the proportion that discontinue an antiretroviral. I shall then focus on two of the toxicities discussed earlier: cardiovascular-related toxicity, via the surrogate marker of total cholesterol; and hepatotoxicity, by considering AST and ALT levels.

5.2 Methods

5.2.1 Patient population

Patients included in analyses are from the Royal Free Hospital Cohort, as described previously. As in Chapter 4, to be included in this study, individuals were required to be previously antiretroviral naïve and starting HAART, defined as a regimen containing at least three antiretrovirals and at least one of a PI, NNRTI or ABA from 1st January 1998 until 31st December 2005. Again, individuals without at least one CD4 cell count measurement and at least one viral load measurement within the six month period prior to starting HAART were excluded.

For the purposes of this Chapter, patients were considered to have HBV if at any time they had a surface antigen test (HBsAg) which was positive or weakly positive, regardless of the timing of the test (that is whether it occurred prior to or after starting HAART). Patients were considered to have HCV if at any time they had a core antibody test (HCVAb) which was positive or weakly positive, regardless of the timing of the test. Pre-HAART CD4 cell count, viral load, total cholesterol, AST and ALT measurements were defined as the result measured nearest to the date of starting HAART, providing it occurred within six months prior to starting HAART. Other characteristics at baseline were summarised for all those starting HAART and in two subgroups. The first

consisted of all those participants with a pre-HAART total cholesterol measurement and at least one follow-up measurement in the period six months to 12 months after starting HAART. The second subgroup contained those with at least one AST or ALT measurement prior to starting HAART, and at least one follow-up measurement in the period six months to 12 months after starting HAART. Thus, I could compare the characteristics of those who would be included in any analyses of these laboratory markers with those who would be excluded.

5.2.2 Impact of the inclusion of dose and formulation switches in the observed proportion of patients discontinuing/switching antiretrovirals in the first year of HAART

As mentioned in the introduction to this section, one potential indicator for the presence of antiretroviral toxicity is discontinuation or switching of an ARV. Therefore, I began by considering the number of individuals who discontinued an ARV during the first year of HAART. I then considered the reasons given for drug discontinuation to investigate whether these were likely to be toxicity related or whether they were likely to be efficacy related. Individuals discontinuing one, two or all of the antiretroviral drugs in their regimen were included. As individuals could stop more than one drug on the same day and could give more than one reason for drug discontinuation, multiple reasons for drug discontinuation were permitted and summarised. As toxicities could be related to the drug levels present in the body (measured approximately by the drug levels present in the plasma) [369] as well as the formulation of the antiretrovirals (e.g. AZT and 3TC administered as two separate drugs rather than in the co-formulation combivir), I first included dose and formulation changes as treatment discontinuations.

To investigate whether ignoring dose/formulation changes impacted on the percentage discontinuing an ARV in the first year of HAART, I reanalysed the data, this time ignoring any dose or formulation changes. To count as a drug discontinuation/switch, individuals had to stop the ARV of interest for at least 2 days, and in these two days they could not re-start the ARV of interest at a different dose or as part of a co-formulation tablet. If an individual stopped a co-formulation tablet and did not re-start it within 2 days, and also did not start both/all components of the tablet as separate ARVs then this was also considered as a treatment discontinuation/switch. This two day margin was incorporated as it was felt that a treatment stop of less than this interval would be unlikely to constitute a treatment interruption. However, a switch from saquinavir hard gel capsules to saquinavir soft gel capsules was not considered a formulation change.

I was also interested in the toxicities associated with the 'third' drug in the regimen, as studies often consider and compare these antiretrovirals specifically [251;306]. Therefore, I repeated the above analyses, but only considering changes and discontinuations to the PI, NNRTI or ABA in the regimen. I only included changes to ABA in this present analysis if it was received without a PI or NNRTI (i.e. it was the 'third' drug in the regimen, and not part of the nucleoside backbone).

5.2.3 Impact of different definitions on the prevalence of antiretroviral-related toxicities observed

I next went on to consider the proportion of individuals who experienced lipid abnormalities, using total cholesterol (TC) levels. I considered three cut-offs in particular: TC >6.2 mmol/l, TC >5.5 mmol/l and a change in TC of >1 mmol/l compared to pre-HAART levels. These cut-offs were chosen as they reflect those used most commonly in the literature, as summarised in Chapter 2 [205;209;251]. Based on my work in Chapter 4, I considered the first measurement in the interval six months to one year after starting HAART. I firstly considered whether the first measurement in the window six months to one year after starting HAART were high (a single high TC measurement), and next considered whether the first two measurements in the window six months to one year after starting HAART were high (a confirmed high TC measurement). All analyses took a missing=excluded approach, and therefore different numbers of individuals were included in analyses as the entry criteria for the different endpoints varied. For example, when considering a single TC change of >1 mmol/l compared to pre-HAART vales as an event, individuals were required to have a pre-HAART TC measurement, as well as at least one TC measurement in the period six months to one year after starting HAART. In contrast, when considering a confirmed TC change of >1 mmol/l compared to pre-HAART values as an event, individuals were required to have a pre-HAART TC measurement, as well as at least two TC measurements in the period six months to one year after starting HAART. I summarised the proportion of individuals starting HAART who experienced a lipid abnormality after 6 months of treatment using the different definitions described above. This enabled me to compare differences in the observed prevalence according to the definition of a toxicity used.

I then went on to investigate how many individuals stopped an ARV during the first year of HAART where the reason given for stopping the drug was a lipid abnormality. I investigated the actual TC levels at the time of the drug discontinuation, to see whether these gave any clues as to what was considered by clinicians and patients as a

clinically important change in TC levels. I also investigated how many of the patients discontinuing an ARV for a lipid abnormality met my definitions of experiencing a TC event.

I next considered the frequency of hepatotoxic events, by considering the liver function (LFT) laboratory markers AST and ALT. I considered the best way to define a toxicity event and the impact of the definition used on the proportion of individuals experiencing an event. I considered the definition of a hepatic toxicity as either an AST/ALT level more than 5 times the upper limit of normal (ULN = 40 IU/L) or an increase of more than 2.5 times the ULN compared to pre-HAART levels, basing these endpoints on those used most commonly in the current literature and summarised in Chapter 2 [302;306]. Again based on my findings in Chapter 4, I considered the first one and first two measurements in the period six months to one year after starting HAART, and investigated the impact of the choice of definition on the proportion thought to be experiencing an event, using a missing=failure approach. I finally investigated how many individuals stopped an ARV during the first year of HAART, where the reason given for stopping the drug was abnormal LFTs. Any ARV stopped during the first year of HAART for the reason of raised LFTs was considered, regardless of whether other ARVs were discontinued at the same time or previously for other reasons. I investigated how many of these individuals were considered to have a hepatotoxicity event based on AST/ALT levels and the definitions above, as well as the actual AST/ALT levels at the time of antiretroviral discontinuation.

5.3 Results: patient characteristics

Of 1281 patients starting HAART between 1st January 1998 and 31st December 2004, 978 (76.3%) patients had a baseline CD4 count and viral load and thus were included in analyses. These patients are described in Table 5.1. The median (inter-quartile range [IQR]) age at starting HAART was 36 (32, 42) years. The majority (719; 73.5%) were male, had a homosexual risk for HIV infection (517; 52.9%) and were of white ethnicity (554; 56.7%). A low prevalence of HBV infection (53; 5.4%) and HCV infection (81; 8.3%) was observed. Two hundred patients (21.6%) did not have an HbsAg test result and thus their HBV status was unknown, although test results before 2001 cannot be considered due to the unreliability of the results. One hundred and fourteen (11.7%) had no HCVAb test result. The number starting HAART was split equally across calendar years; 249 (25.5%) started HAART in 1998-1999, 418 started in 2000-

2002 (42.7%) and 311 started in 2003-2004 (31.8%). The median (IQR) CD4 cell count and viral load at the time of starting HAART were 189 (77, 295) cells/mm^3 and 5.2 (4.7, 5.7) log copies/ml respectively. The most common starting regimen was NNRTI+2NRTI (490; 50.1%), followed by PI+RTV+2NRTI (259; 26.5%) and PI+2NRTI (129; 13.2%).

Also shown in Table 5.1 are the characteristics of those with a pre-HAART TC measurement, and at least one follow-up measurement in the period six months to one year after starting HAART. Four hundred and eighty (49.0% of those who started HAART) could be included. Compared to all those starting HAART, those included in the TC subgroup had a similar age, gender, ethnicity, risk group, HCV, HBV, CD4 cell count and viral load distribution. However, only 60 (12.5%) started HAART in 1998-1999, compared to 249 (25.5%) of all those starting HAART. The most common regimen type was again NNRTI+2NRTI (228; 47.5%), but PI+RTV+2NRTI regimens were much more common (173; 36.0%) than in the entire cohort. The median (IQR) TC at the start of HAART was 4.0 (3.5, 4.7) mmol/l, 27 (5.6%) had a TC >5.5 mmol/l and 12 (2.5%) had a TC >6.2 mmol/l.

Six hundred and seventy patients (68.5% of those who started HAART) had a pre-HAART AST/ALT measurement and at least one follow-up measurement in the period six months to one year after starting HAART. All characteristics were similar to those seen in all 978 patients starting HAART. The median (IQR) AST and ALT measurement at HAART was 31 (24, 41) IU/L and 30 (20, 45) IU/L respectively. One Two hundred and forty one (36.0%) had an AST or ALT measurement above the ULN at baseline (40 IU/L).

Table 5.1 – Characteristics of patients included in present analyses at the time of starting HAART

		All patients	TC subgroup*	AST/ALT subgroup*
Number		978 (100.0%)	480 (100.0%)	670 (100.0)
Age (years)	Median (IQR)	36 (32, 42)	37 (32, 42)	36 (32, 41)
Gender	Male	719 (73.5)	361 (75.2)	486 (72.5)
Risk group	Homosexual	517 (52.9)	271 (56.5)	364 (54.3)
	Heterosexual	417 (42.6)	193 (40.2)	279 (41.6)
	Other	44 (4.5)	16 (3.3)	27 (4.0)
Ethnicity	White	554 (56.7)	280 (58.3)	395 (59.0)
	Black African	292 (29.9)	148 (30.8)	202 (30.2)
	Other	132 (13.5)	52 (10.8)	73 (10.9)
HBV positive	Yes	53 (5.4)	30 (6.3)	39 (6.9)
	No	725 (74.1)	387 (86.0)	528 (83.7)
	Unknown	200 (21.6)	63 (14.0)	103 (16.3)
HCV positive	Yes	81 (8.3)	37 (7.7)	60 (9.0)
	No	783 (80.1)	413 (93.2)	566 (92.8)
	Unknown	114 (11.7)	30 (6.8)	44 (7.2)
Calendar year	1998-1999	249 (25.5)	60 (12.5)	149 (22.2)
	2000-2002	418 (42.7)	186 (38.8)	282 (42.1)
	2003-2004	311 (31.8)	234 (48.8)	239 (35.7)
CD4 cell count (cells/mm^3)	Median (IQR)	189 (77, 295)	199 (94, 298)	195 (89, 294)
Viral load (log copies/ml)	Median (IQR)	5.2 (4.7, 5.7)	5.2 (4.7, 5.7)	5.2 (4.8, 5.7)
Regimen type	PI+2NRTI	129 (13.2)	38 (7.9)	79 (11.8)
	NNRTI+2NRTI	490 (50.1)	228 (47.5)	338 (50.5)
	PI+RTV+2NRTI	259 (26.5)	173 (36.0)	196 (29.3)
	Other	100 (10.2)	41 (8.5)	57 (8.5)
NRTIs	AZT+3TC	513 (52.5)	237 (49.4)	349 (52.1)
	ddI+d4T	56 (5.7)	23 (4.8)	36 (5.4)
	TDF+FTC	26 (2.7)	19 (4.0)	19 (2.8)
	TDF+3TC	57 (5.8)	43 (9.0)	45 (6.7)
	3TC+d4T	178 (18.2)	83 (17.3)	124 (18.5)
	AZT+3TC+ABA	34 (3.5)	10 (2.1)	19 (2.8)
	Other	114 (11.7)	65 (13.5)	78 (11.6)
Third drug	NVP	152 (15.5)	47 (9.8)	91 (13.6)
	EFV	366 (37.4)	190 (39.6)	262 (39.1)
	ABA	25 (2.6)	11 (2.3)	15 (2.2)
	LPV/r	208 (21.3)	140 (29.2)	154 (23.0)
	IDV	31 (3.2)	8 (1.7)	20 (3.0)
	NFV	80 (8.2)	25 (5.2)	47 (7.0)
	Other	116 (11.9)	59 (12.3)	81 (12.1)
TC (mmol/l)	Median (IQR)	-	4.0 (3.5, 4.7)	-
TC>5.5 mmol/l	Yes	-	27 (5.6)	-
TC>6.2 mmol/l	Yes	-	12 (2.5)	-
AST (IU/L)	Median (IQR)	-	-	31 (24, 41)
ALT (IU/L)	Median (IQR)	-	-	30 (20, 45)
AST>40 IU/L	Yes	-	-	177 (26.4)
ALT >40 IU/L	Yes	-	-	195 (29.1)
AST or ALT >40 IU/L	Yes	-	-	241 (36.0)

TC=total cholesterol; * TC subgroup: Patients included had a pre-HAART total cholesterol measurement and at least 1 measurement 6-12 months after starting HAART. AST/ALT subgroup: Patients included had a pre-HAART AST/ALT measurement and at least 1 measurement 6-12 months after starting HAART

5.4 Results: Antiretroviral discontinuation

In the first year of HAART, 483 individuals (49.4%) made at least one change to their antiretroviral regimen when treating dose and formulation changes as treatment changes. The reasons for the first switch made to the regimen were varied (Table 5.2; more than one reason for drug discontinuation or switch was permitted. A total of 34 (7.0% of those making a change to the regimen) did so for reasons relating to treatment failure; 29 gave a reason for stopping as viral load failure (6.0%), 1 (0.2%) as CD4 failure and 6 (1.2%) as increased drug resistance. A total of 217 (44.9%) gave a toxicity reason for discontinuing or stopping an ARV. The most common toxicities reported were nausea and vomiting (37; 7.7%), anaemia (40; 8.3%), peripheral neuropathy (27; 5.6%) and CNS effects/insomnia (43; 8.9%). Two hundred and seventy three (56.5%) gave other reasons for discontinuing an ARV. The most common reasons given were patient choice (71; 14.7%) and rationalisation (82; 17.0%).

When excluding dose changes and formulation changes, 434 (44.4%) of those starting HAART discontinued or switched an antiretroviral in the first year of HAART (Table 5.2). In 33 (7.6%) of cases one of the reasons for ARV discontinuation was drug failure. Two hundred and thirteen (49.1%) patients first made a regimen change citing a toxicity-related reason. The most common were nausea/vomiting (34; 7.8%), anaemia (35; 8.1%), peripheral neuropathy (29; 6.7%), CNS effects/insomnia (46; 10.6%) and rash 25 (5.8%). Other reasons were given in 214 (49.3%) cases, with patient choice begin cited in 76 (17.5%) cases, and rationalization being much less common than when dose/formulation changes were included at 27 (6.2%) cases.

Table 5.2 – Percentage discontinuing or switching at least one ARV in the first year of HAART and reasons given for first treatment change

	Including dose /formulation changes	**Excluding dose/ formulation changes**
Total number of patients	978	978
Number discontinuing an ARV	483 (100.0%)	434 (100.0%)
Treatment failure		
Any	34 (7.0%)	33 (7.6%)
failure – viral load	29 (6.0%)	28 (6.5%)
failure CD4	1 (0.2%)	1 (0.2%)
Failure, increased resistance	6 (1.2%)	6 (1.4%)
Toxicity		
Any	217 (44.9%)	213 (49.1%)
renal problem	4 (0.8%)	4 (0.9%)
nausea/vomiting	37 (7.7%)	34 (7.8%)
anaemia	40 (8.3%)	35 (8.1%)
lipodystrophy	2 (0.4%)	2 (0.5%)
abdominal pain	1 (0.2%)	1 (0.2%)
malaise/fatigue	4 (0.8%)	4 (0.9%)
GI side effects	4 (0.8%)	4 (0.9%)
peripheral neuropathy	27 (5.6%)	29 (6.7%)
CNS effects/insomnia	43 (8.9%)	46 (10.6%)
depression	0 (0.0%)	0 (0.0%)
abnormal LFTs	9 (1.9%)	8 (1.8%)
intolerance	0 (0.0%)	0 (0.0%)
rash	25 (5.2%)	25 (5.8%)
allergic reaction	8 (1.6%)	8 (1.8%)
diarrhoea	15 (3.1%)	15 (3.5%)
lipid abnormality	2 (0.4%)	2 (0.5%)
pancreatitis	2 (0.4%)	2 (0.5%)
headache	3 (0.6%)	3 (0.7%)
lactic acidosis	1 (0.2%)	0 (0.0%)
fat wasting	1 (0.2%)	1 (0.2%)
fat accumulation	3 (0.6%)	3 (0.7%)
Other		
Any	273 (56.5%)	214 (49.3%)
tetragenic effects in pregnancy	0 (0.0%)	0 (0.0%)
only to prevent vertical transmission	12 (2.5%)	13 (3.0%)
Study endpoint	29 (6.0%)	30 (6.9%)
drug interaction	8 (1.7%)	4 (0.9%)
following a TDM result	1 (0.2%)	0 (0.0%)
rationalisation	82 (17.0%)	27 (6.2%)
poor compliance	12 (2.5%)	12 (2.8%)
patient choice	71 (14.7%)	76 (17.5%)
other	43 (8.3%)	34 (7.8%)
not recorded in notes	28 (5.8%)	22 (5.1%)

Three hundred and fifty one (35.9%) patients had made a change to the third drug (PI or NNRTI or ABA) in their regimen in the first year of HAART when including dose and formulation changes (Table 5.3). In 41 (11.7%) cases treatment failure was given as a reason for treatment discontinuation/switch and a toxicity reason was given in 135 (38.5%) of cases. The most common toxicities reported were nausea and vomiting (25; 7.1%), CNS effects and insomnia (49; 14.0%) and rash (23; 6.6%). In contrast to the situation in which discontinuation of any antiretroviral were included, no patient gave peripheral neuropathy as a reason for stopping. Other reasons were given in 192 (54.7%) of treatment changes; 40 (11.4%) were due to rationalization and 55 (15.7%) for patient choice.

When excluding dose and formulation changes, approximately a third (312; 31.9%) of the 978 patients made a change to their PI/NNRTI in their first year of HAART. In 35 (11.2%) cases, a reason given for this switch was related to treatment failure. In 132 (42.3%) of cases, it was toxicity related; 22 (7.1%) cases gave nausea/vomiting as a reason for treatment change, 49 (15.7%) cases were CNS effects/insomnia and 25 (8.0%) for rash. Again, no changes were made for peripheral neuropathy. One hundred and fifty four (49.4%) treatment changes were made for other reasons, including 57 (18.3%) changes for patient choice. Again, the number switching for rationalisation (16; 5.1% of changes) was much reduced compared to the situation when including dose and formulation changes.

Table 5.3 – Percentage discontinuing or switching the third drug in the first year of HAART and reasons for first treatment change

	Including dose /formulation changes	**Excluding dose/ formulation changes**
Total number of patients	978	978
Number discontinuing a PI/NNRTI	351 (100.0%)	312 (100.0%)
Treatment failure		
Any	41 (11.7%)	35 (11.2%)
failure – viral load	35 (10.0%)	29 (9.3%)
failure CD4	1 (0.3%)	1 (0.3%)
Failure, increased resistance	7 (2.0%)	7 (2.2%)
Toxicity		
Any	135 (38.5%)	132 (42.3%)
renal problem	4 (1.1%)	2 (0.6%)
nausea/vomiting	25 (7.1%)	22 (7.1%)
anaemia	2 (0.6%)	1 (0.3%)
lipodystrophy	1 (0.3%)	1 (0.3%)
abdominal pain	1 (0.3%)	1 (0.3%)
malaise/fatigue	3 (0.9%)	3 (1.0%)
GI side effects	3 (0.9%)	3 (1.0%)
peripheral neuropathy	0 (0.0%)	0 (0.0%)
CNS effects/insomnia	49 (14.0%)	49 (15.7%)
abnormal LFTs	8 (2.3%)	8 (2.6%)
intolerance	0 (0.0%)	0 (0.0%)
rash	23 (6.6%)	25 (8.0%)
allergic reaction	5 (1.4%)	5 (1.6%)
diarrhoea	13 (3.7%)	15 (4.8%)
lipid abnormality	3 (0.9%)	3 (1.0%)
pancreatitis	0 (0.0%)	0 (0.0%)
headache	0 (0.0%)	0 (0.0%)
lactic acidosis	0 (0.0%)	0 (0.0%)
fat wasting	0 (0.0%)	0 (0.0%)
fat accumulation	1 (0.3%)	1 (0.3%)
Other		
Any	192 (54.7%)	154 (49.4%)
tetragenic effects in pregnancy	0 (0.0%)	0 (0.0%)
only to prevent vertical transmission	12 (3.4%)	13 (4.2%)
Study endpoint	22 (6.3%)	20 (6.4%)
drug interaction	6 (1.7%)	2 (0.6%)
following a TDM result	3 (0.9%)	0 (0.0%)
rationalisation	40 (11.4%)	16 (5.1%)
poor compliance	8 (2.3%)	8 (2.6%)
patient choice	55 (15.7%)	57 (18.3%)
other	30 (8.6%)	24 (7.7%)
not recorded in notes	18 (5.1%)	14 (4.5%)

5.5 Results: total cholesterol (TC) events

In total 582 (59.5%) of the 978 individuals starting HAART had a pre-HAART total cholesterol measurement. During the first year of HAART there was a total of 3886 TC measurements recorded, with a median (range) of 4 (0, 20) measurements per individual. One hundred and seventy eight individuals (18.2%) had no TC measurements, 102 (10.4%) had one measurement only, 101 (10.3%) had two measurements, 94 (9.6%) had three measurements, 87 (8.9%) had four measurements, 108 (11.0%) had five measurements, 111 (11.4%) had six measurements and 197 (20.1%) had seven or more measurements. Six hundred and eighty five individuals (70.0%) had at least one TC measurement in the period six months to one year after starting HAART, and 468 (47.9%) had two or more measurements in this period. A total of 480 patients (49.1%) had both a pre-HAART measurement and at least one measurement six months to one year after starting HAART (these patients are described in Table 5.1) and 358 (36.6%) had both a pre-HAART measurement and at least two measurements six months to one year after starting HAART.

The number of individuals considered to have experienced a TC event in the first year of HAART was strongly affected by the definition used, with between 4.2% and 36.0% considered to have experienced a TC event (Table 5.4). I firstly considered the 480 patients with a pre-HAART TC and at least one follow-up measurement in the period six months to one year after starting HAART. Eighty patients' (16.7%) first measurement in that period was >5.5 mmol/l and also >1 mmol/l higher than pre-HAART levels (definition 1); 39 (8.1%) experienced a first TC >6.2 mmol/l during this period HAART that was also >1 mmol/l higher than pre-HAART levels (definition 2). Amongst the subgroup of 358 patients with a pre-HAART TC and at least 2 measurements in the period six months to one year after starting HAART, 41 (11.5%) had a first and second measurement in the time interval that were both >5.5 mmol/l and also >1 mmol/l higher than the pre-HAART TC (definition 3) and 15 (4.2%) had a confirmed TC >6.2 mmol/l which was also >1 mmol/l higher than the pre-HAART TC (definition 4).

My next definitions excluded those who had high TC levels before starting HAART as these endpoints only considered the absolute value of TC levels and not the change from pre-HAART levels. When excluding those with a pre-HAART TC >5.5 mmol/l, 99/453 (21.9%) had a first TC>5.5 mmol/l in the period six months to one year after starting HAART (definition 5) and 55/336 (16.4%) had a confirmed TC>5.5 mmol/l (definition 7). Forty two of 468 (9.0%) of those with a pre-HAART TC ≤6.2 mmol/l had

at least one TC>6.2 mmol/l in the period six months to one year after starting HAART (definition 6) and 15/347 (4.3%) had a confirmed TC>6.2 mmol/l (definition 8).

I next considered changes in TC from pre-HAART without considering the absolute value. Of 480 patients, 173 (36.0%) experienced an increase in TC of >1 mmol/l compared to pre-HAART levels on at least one occasion in the period six months to one year after starting HAART (definition 9), and 97/358 (27.1%) experienced a confirmed >1 mmol/l increase in TC (definition 10).

My next definitions were comparable to definitions 5-8, but included all individuals with follow-up TCs, regardless of the presence of a pre-HAART measurement. One hundred and ninety one of 685 individuals (27.9%) had a first TC>5.5 mmol/l in the period six months to one year after starting HAART (definition 11) and 96/468 (20.5%) had a confirmed TC>5.5 mmol/l in this period (definition 13). In total 79/685 (11.5%) had at least one TC>6.2 mmol/l (definition 12) and 32/468 (6.8%) had a confirmed TC>6.2 mmol/l (definition 14).

Table 5.4 – Percentage experiencing lipid abnormality according to definition used

	Required TC measurements for inclusion					Definition of event	Number included	Number (%) with event
		Pre-HAART		Number of measurements in period 6-12 months after HAART				
	Any	≤5.5 mmol/l	≤6.2 mmol/l	At least 1	At least 2			
1	X			X		(1) Single TC>5.5 mmol/l (2) Measurement also >1 mmol/l higher than pre-HAART	480	80 (16.7%)
2	X			X		(1) Single TC>6.2 mmol/l (2) Measurement also >1 mmol/l higher than pre-HAART	480	39 (8.1%)
3	X				X	(1) Confirmed TC>5.5 mmol/l (2) Measurements also >1 mmol/l higher than pre-HAART	358	41 (11.5%)
4	X				X	(1) Confirmed TC>6.2 mmol/l (2) Measurements also >1 mmol/l higher than pre-HAART	358	15 (4.2%)
5		X		X		(1) Single TC >5.5 mmol/l	453	99 (21.9%)
6			X	X		(1) Single TC >6.2 mmol/l	468	42 (9.0%)
7		X			X	(1) Confirmed TC >5.5 mmol/l	336	55 (16.4%)
8			X		X	(1) Confirmed TC >6.2 mmol/l	347	15 (4.3%)
9	X			X		(1) Single TC >1 mmol/l higher than pre-HAART	480	173 (36.0%)
10	X				X	(1) Confirmed TC >1 mmol/l higher than pre-HAART	358	97 (27.1%)
11				X		(1) Single TC>5.5 mmol/l	685	191 (27.9%)
12				X		(1) Single TC>6.2 mmol/l	685	79 (11.5%)
13					X	(1) Confirmed TC>5.5 mmol/l	468	96 (20.5%)
14					X	(1) Confirmed TC>6.2 mmol/l	468	32 (6.8%)

TC=total cholesterol; HAART=highly active antiretroviral therapy; confirmed=two consecutive measurements.

During the first year of HAART only three patients stopped an antiretroviral citing a lipid abnormality as a reason for discontinuation. In two cases the antiretroviral discontinued was lopinavir and in one case the drug was ritonavir (which was being administered as the sole PI in the regimen). The pre-HAART TC levels were 2.9 mmol/l, 4.5 mmol/l and at the time of discontinuing the drug these had become 5.7 mmol/l, 10 mmol/ and 8.3 mmol/l respectively. I next considered whether the individuals stopping an ARV for raised lipids would meet the definition of experiencing an event as defined in Table 5.4. Any definitions in which a pre-HAART TC ≤5.5 mmol/l or ≤6.2 mmol/l was required (definitions 5-8) meant that one individual was excluded. However, all three had a pre-HAART TC measurement and at least 2 TC measurements in the period six months to one year after starting HAART. Thus, as can be seen in Table 5.5, even though numbers were small, in a number of analyses not all individuals were considered as having experienced a TC event, even though the level of the TC was thought to be clinically significant and require intervention. No definition captured all three patients who changed treatment for this reason.

Table 5.5 – Patients discontinuing an ARV in the first year of HAART where the reason given for discontinuation was lipid abnormality – proportions experiencing total cholesterol endpoints as defined in Table 5.4

Definition of TC event (see Table 5.4)		Excluded	No event	Event
1	(1) Single TC>5.5 mmol/l (2) Also >1 mmol/l higher than pre-HAART	0 (0.0%)	2 (66.7%)	1 (33.3%)
2	(1) Single TC>6.2 mmol/l (2) Also >1 mmol/l higher than pre-HAART	0 (0.0%)	3 (100.0%)	0 (0.0%)
3	(1) Confirmed TC>5.5 mmol/l (2) Also >1 mmol/l higher than pre-HAART	0 (0.0%)	2 (66.7%)	1 (33.3%)
4	(1) Confirmed TC>6.2 mmol/l (2) Also >1 mmol/l higher than pre-HAART	0 (0.0%)	3 (100.0%)	0 (0.0%)
5	(1) Single TC >5.5 mmol/l	1 (33.3%)	1 (33.3%)	1 (33.3%)
6	(1) Single TC >6.2 mmol/l	1 (33.3%)	2 (66.7%)	0 (0.0%)
7	(1) Confirmed TC >5.5 mmol/l	1 (33.3%)	1 (33.3%)	1 (33.3%)
8	(1) Confirmed TC >6.2 mmol/l	1 (33.3%)	2 (66.7%)	0 (0.0%)
9	(1) Single TC >1 mmol/l higher than pre-HAART TC	0 (0.0%)	1 (33.3%)	2 (66.7%)
10	(1) Confirmed TC >1 mmol/l higher than pre-HAART	0 (0.0%)	1 (33.3%)	2 (66.7%)
11	(1) Single TC>5.5 mmol/l	0 (0.0%)	1 (33.3%)	2 (66.7%)
12	(1) Single TC>6.2 mmol/l	0 (0.0%)	2 (66.7%)	1 (33.3%)
13	(1) Confirmed TC>5.5 mmol/l	0 (0.0%)	1 (33.3%)	2 (66.7%)
14	(1) Confirmed TC>6.2 mmol/l	0 (0.0%)	2 (66.7%)	1 (33.3%)

5.6 Results: hepatotoxic events

I next considered the occurrence of hepatotoxicity events. Of 978 individuals starting HAART, 771 (78.8%) had a pre-HAART AST measurement, and 771 (78.8%) had a pre-HAART ALT measurement. There was a total of 5782 AST and 5781 ALT measurements in the first year of HAART with a median (IQR) of 5 (0, 61) and 5 (0, 61) AST and ALT measurements per person respectively. Seven hundred and seventy two (78.9%) had at least one AST measurement in the period six months to one year after starting HAART, and 600 (61.3%) had at least two measurements; the same number had ALT measurements.

Again, the percentage experiencing an LFT event varied depending on the definition used (Table 5.6), ranging from 0.2% to 1.3%. However, regardless of the definition used, few hepatotoxicity events occurred. Six hundred and seventy individuals had a pre-HAART AST/ALT measurement and at least one measurement in the period six months to one year after starting HAART. Of these, 6 individuals (0.9%) had a first measurement in this window that was >5 times the ULN and also >2.5 times the ULN higher than pre-HAART levels (definition 1). Two of 522 (0.4%) had a confirmed AST/ALT >5 times the ULN with values that were also >2.5 times the ULN higher than pre-HAART levels (definition 2). When only considering those with pre-HAART AST and ALT levels below the ULN (<40 IU/L), three patients out of 539 (0.6%) had at least one AST/ALT >5 times the ULN (definition 3) and one patient out of 428 (0.2%) had a confirmed AST/ALT >5 times ULN (definition 4). Again considering the 670 individuals with a pre-HAART AST/ALT measurement and at least one follow-up measurement, 9 (1.3%) had an increase in AST/ALT of >2.5 times ULN (definition 5); 2/522 (0.4%) had a confirmed increase in AST/ALT of >2.5 times ULN (definition 6).

There were a total of thirteen patients who discontinued an ARV (not necessarily their first antiretroviral discontinuation) in the first year of HAART for the reason of abnormal LFTs. The characteristics of these patients are shown in Table 5.7. There was a variety of different ARVs discontinued for the reason of hepatotoxicity, the most common being d4T (4; 30.8%). Two (15.4%) were HCV positive and five were HBV positive (38.5%). The pre-HAART AST and ALT was unknown in 3 (23.1%) of individuals; the median (range) values were 35 (24, 159) IU/L and 45 (18, 167) IU/L respectively. At the time of ARV discontinuation the median (range) AST and ALT levels were 174 (57, 491) and 194 (47, 953) IU/L respectively; the values were missing for 2 patients. Just 4 (30.8%) patients had an AST >5 times ULN at treatment discontinuation; 5 (38.5%) had an ALT >5 times ULN.

Table 5.6 – Percentage experiencing hepatic event according to definition of event used

Required AST/ALT measurements to be included					Definition of event	Number included	Number with event
	Pre-HAART measurement		Number of measurements in 6-12 months after HAAART				
	AST and/ or ALT	(1) AST and/or ALT (2) if present then <40 IU/L	At least 1	At least 2			
1	X		X		(1) Single ALT>5 ULN or single AST >5 ULN (2) Measurement is also >2.5 ULN higher than pre-HAART AST/ALT	670	6 (0.9%)
2	X			X	(1) Confirmed ALT>5 ULN or confirmed AST>5 ULN (2) Measurements are also >2.5 ULN higher than pre-HAART AST/ALT	522	2 (0.4%)
3		X	X		(1) Single ALT>5 ULN or single AST >5 ULN	539	3 (0.6%)
4		X		X	(1) Confirmed ALT>5 ULN or confirmed AST>5 ULN	428	1 (0.2%)
5	X		X		(1) Single AST >2.5 ULN higher than pre-HAART AST or single ALT >2.5 ULN higher than pre-HAART ALT	670	9 (1.3%)
6	X			X	(1) Confirmed AST >2.5 ULN higher than pre-HAART AST or confirmed ALT >2.5 ULN higher than pre-HAART ALT	522	2 (0.4%)

Table 5.7 – Characteristics of those stopping an ARV for the reason of abnormal LFTs in first year of HAART

		Number (%)
Number		13 (100.0%)
ARV stopped	AZT	1 (7.7%)
	ddI	1 (7.7%)
	d4T	4 (30.8%)
	AZT/3TC	2 (15.4%)
	FTC	1 (7.7%)
	NVP	1 (7.7%)
	EFV	2 (15.4%)
	LPV	1 (7.7%)
HCV status	Positive	2 (15.4%)
	Negative	11 (84.6%)
HBV status	Positive	5 (38.5%)
	Negative	8 (61.5%)
Pre –HAART AST (IU/L)	Missing	3 (23.1%)
	Median (range)	35 (24, 159)
Pre –HAART ALT (IU/L)	Missing	3 (23.1%)
	Median (range)	45 (18, 167)
AST at time of discontinuation (IU/L)	Missing	2 (15.4%)
	Median (range)	174 (57, 491)
ALT at time of discontinuation (IU/L)	Missing	2 (15.4%)
	Median (range)	194 (47, 953)
AST at discontinuation>200 IU/L	Yes	4 (30.8%)
ALT at discontinuation>200 IU/L	Yes	5 (38.5%)
2 previous AST at discontinuation>200 IU/L	Yes	2 (15.4%)
2 previous ALT at discontinuation>200 IU/L	Yes	1 (7.7%)
AST change from pre-HAART at ARV discontinuation (IU/L)	Missing	3 (23.1%)
	Median (range)	+138 (0, +455)
ALT change from pre-HAART at ARV discontinuation (IU/L)	Missing	3 (23.1%)
	Median (range)	+121 (-4, +836)
First AST measurement in period 6 months to one year after starting HAART (IU/L)	Missing	2 (15.4%)
	Median (range)	97 (13, 2120)
First ALT measurement in period 6 months to one year after starting HAART (IU/L)	Missing	2 (15.4%)
	Median (range)	60 (16, 962)

A small proportion of the individuals stopping an ARV for the reason of abnormal LFTs met the definition of an LFT event as defined in Table 5.6 (results shown in Table 5.8). Using definition 1 (at least one AST/ALT >5 times ULN during first year and measurement also >2.5 times ULN than pre-HAART value) four individuals were excluded, 6 did not meet the definition of an event and 3 were considered to have an event. With definitions two, three and four, one (7.7%), two (15.4%) and none (0.0%) of those individuals who discontinued an ARV for abnormal LFTs met the definition of an LFT event. The highest percentage of individuals were considered to have an LFT event when considering definition 5, which was at least one AST/ALT measurement in the first year of HAART of >2.5 ULN higher than pre-HAART values; here 4 (30.8%) patients were considered to have an event and 4 (30.8%) were excluded from analyses; when a confirmed event was required (definition 6), one patient still met this definition (7.7%).

Table 5.8– Characteristics of patients discontinuing an ARV in the first year of HAART where the reason given is abnormal LFTs

	Definition of LFT event number (Table 5.7)	Excluded	No event	Event
1	(1) Single ALT>5 times ULN or single AST >5 times ULN (2) Measurement is also >2.5 times ULN higher than pre-HAART AST/ALT	4 (30.8%)	6 (46.2%)	3 (23.1%)
2	(1) Confirmed ALT>5 times ULN or confirmed AST>5 times ULN (2) Measurements are also >2.5 times ULN higher than pre-HAART AST/ALT	5 (38.5%)	7 (53.9%)	1 (7.7%)
3	(1) Single ALT>5 times ULN or single AST >5 times ULN	6 (46.2%)	5 (38.5%)	2 (15.4%)
4	(1) Confirmed ALT>5 times ULN or confirmed AST>5 ULN	7 (53.9%)	6 (46.2%)	0 (0.0%)
5	(1) Single AST >2.5 times ULN higher than pre-HAART AST or single ALT >2.5 times ULN higher than pre-HAART ALT	4 (30.8%)	5 (38.5%)	4 (30.8%)
6	(1) Confirmed AST >2.5 ULN times higher than pre-HAART AST or confirmed ALT >2.5 times ULN higher than pre-HAART ALT	5 (38.5%)	7 (53.9%)	1 (7.7%)

5.7 Discussion

As the Royal Free Hospital database only collects information on the clinical diagnosis of toxicities which lead to death, admission as an in-patient, or those that are easily captured by a surrogate laboratory marker. Other toxicities not captured by the above cannot be easily studied. One possible option is to investigate those who are unable to tolerate antiretrovirals, and thus stop or make changes to their antiretroviral regimen. As we have seen in this chapter, a large number of patients starting first-line HAART at the Royal Free Hospital make changes to their regimen in the first year of HAART, perhaps indicating that there is a high prevalence of drug-related toxicity in this cohort.

However, there are limitations to this approach. Firstly, the level to which an individual is able to tolerate a toxicity such as nausea will be subjective and will vary from patient to patient. Secondly, there is the issue of competing risks [370]: those who first discontinue or switch a drug because of lack of efficacy are unable to be considered as first stopping a drug for toxicity reasons. This issue is addressed in further detail in Chapter 6. Thirdly, clinical efficacy and toxicity may not be mutually exclusive events; an individual may find an antiretroviral difficult to tolerate which could potentially lead to poor adherence and treatment failure. This may be recorded as stopping the drug for failure when the underlying cause was also toxicity-related, and thus may again not be captured in these analyses. Furthermore, those experiencing toxicities but who do not stop or change their antiretroviral regimen are not captured in this endpoint. Finally, it is difficult to know whether 'other' causes for treatment change are toxicity-related; for example it is not clear whether stopping for "patient choice" should be counted as a toxicity-related treatment discontinuation.

I have found that the inclusion of dose and formulation changes as a treatment change did not lead to substantially different results compared to when dose and formulation changes were excluded. 49.4% discontinued at least one ARV in the first year of HAART when including these events and 44.4% did so when excluding them; 35.9% discontinued the third drug in the regimen when including dose and formulation changes and 31.9% did so when these events were excluded. Therefore, a high proportion of patients discontinued antiretrovirals in their regimen, regardless of the definition applied. Additionally, the proportion reporting most reasons for stopping treatment remained consistent regardless of how dose/formulation changes were incorporated. The one exception is rationalisation. This endpoint is likely to capture changes from two separate tablets to co-formulation tablets, and thus the proportion citing this as a reason for discontinuation was much lower when excluding dose/formulation changes. Dose changes may not always be toxicity related,

particularly if the dose of a drug is increased, as this is likely to be associated with plasma drug levels being too low for the drug to be clinically effective. On the other hand, excess drug levels may be associated with increased levels of toxicity [360], and therefore one may wish to capture these changes. One potential solution may be to include information on virological outcome in the definition of an event (i.e. those discontinuing an ARV with a viral load <50 copies/ml are assumed to be stopping for toxicity reasons), although even after incorporating this extra information one may not be completely sure of the reason for treatment discontinuation. Nonetheless, in the investigations carried out in this chapter at least, it appears as if the decision to include or exclude these treatment changes is not likely to impact greatly on any results observed.

In this chapter I have considered the prevalence of antiretroviral toxicity when considering all HAART regimens, and have not compared different regimens. Although the principles discussed here are likely to be similar, regardless of the particular antiretrovirals investigated, there may be some additional issues to consider. As an example, consider a study investigating whether the proportion discontinuing the PI lopinavir/ ritonavir and efavirenz in the first year of treatment was comparable in which dose and formulation changes were considered a treatment change. In this situation, a new LPV/r tablet is currently being introduced to replace the soft gel capsule [371]. Patients who at the moment are receiving the soft gel capsule formulation are likely to be switched to the new formulation, and this switch is therefore unlikely to be toxicity related. Therefore, any analysis including dose formulations in this situation is likely to be biased.

It is clear that the definition of a total cholesterol event and an LFT event can impact on the prevalence of antiretroviral-related toxicity observed. Although the percentage experiencing an event is likely to be influenced by those who can be included in each analysis, the definition used also impacts on the prevalence observed. It is not immediately obvious which is the most appropriate to use; one must weigh up the sensitivity and specificity of the events (i.e. we wish to have a definition of an event which is sensitive enough to capture those who experience the event, but is specific enough to rule out an event in those who are not truly at risk of an event). Considering the clinical implications of an event occurring may help us to decide which is the most relevant endpoint to use.

In the general population, total cholesterol levels are linearly associated with an increased risk of cardiovascular disease [372]; an increase from 3 mmol/l to 4 mmol/l in

total cholesterol is associated with the same increase in risk of cardiovascular disease as an increase from 7 mmol/l to 8 mmol/l. Therefore, any increase in total cholesterol levels that may be associated with antiretroviral treatment is likely to be detrimental to the patient. This suggests that considering the change in total cholesterol levels, such as definitions 9 and 10 in Table 5.4, may be particularly appropriate. Conversely, the absolute risk of cardiovascular disease becomes significant at higher levels, and if the aim is to identify those at a high risk of cardiovascular disease it may be appropriate to only consider those with high total cholesterol levels. At the time I began my thesis, the Royal Free Hospital defined an abnormal total cholesterol as >6.2 mmol/l, as have several studies in the HIV setting [209;218;251;253], but current UK Department of Health guidelines suggest that individuals' TC should be maintained lower than 5.0 mmol/l [373]. However, it is also important to consider the patient population being studied here; there are a large proportion of men, the average age is 35 years, and it has been shown that there is a high prevalence of smoking [254], all of which are risk factors for cardiovascular disease [363]. Thus, perhaps a more conservative cut-off is more appropriate. Even here we must consider how much of the effect is due to the treatment itself, and how much to lifestyle factors. Therefore, perhaps a composite endpoint incorporating changes in total cholesterol levels as well as the absolute value is most appropriate.

Exactly who should be included in these analyses is also important. Clearly it would be preferable to include as many individuals as possible to minimise any selection biases, and thus removal of the restriction on the presence of a pre-HAART total cholesterol measurement would be an attractive option. This would also serve to increase the power of any analyses as the number of patients included would be higher. However, we wish to be able to distinguish between increases in total cholesterol observed whilst an individual is receiving HAART and the persistence of high TC values that are present before starting HAART. We could also consider whether a single high value during HAART or a confirmed (two consecutive values) high TC is the most appropriate endpoint. Clearly, if only one value is required the number of individuals who can be included in the analysis increases. Conversely, as there is a certain amount of natural variability and measurement error in total cholesterol measurements [374], and the samples analysed here are both fasting and non-fasting, it may be preferable to consider a confirmed result. Certainly adding this restriction reduced the percentage experiencing an event compared to a single high value in the analyses presented in Table 5.4, suggesting that some of the total cholesterol increases may have been transient. All of the above points may perhaps suggest that a reasonable endpoint to use to define a total cholesterol event is definition 3; a confirmed TC>5.5 mmol/l which

is also >1 mmol/l higher then pre-HAART TC. Using this definition, we observed hypercholesterolaemia in 11.5% of our population (41/358 people).

There are limitations when considering a total cholesterol endpoint as a surrogate marker at all in our analysis. Firstly, we have not considered the other cholesterol markers (triglycerides, low density lipoprotein (LDL) cholesterol and high density lipoprotein (HDL)). Clearly it is also important to know whether total cholesterol levels are increasing as a result of an increase in HDL cholesterol (and therefore beneficial to the patient), or because of an increase in triglycerides and LDL cholesterol levels. Although these sub-fractions have not been discussed in detail here, the principles applied in this Chapter for these other cardiovascular markers apply equally. Lipid levels only constitute one factor contributing to the risk of cardiovascular disease [372] and others, such as blood pressure and family history, cannot be studied here. Secondly, I have not considered the impact of other interventions for cardiovascular disease, such as statins and dietary interventions. Clearly, if changes in total cholesterol are prompting the treating physician to intervene in a way other than changing the antiretrovirals received and before the patient meets the definition for hypercholesterolaemia then they would not be captured in our endpoints. Unfortunately it is not easy to capture these data succinctly. Although we have information on those patients in the Royal Free Hospital who are referred to the specialist HIV lipids clinic, set up to offer advice to those at a high risk of cardiovascular disease (or as secondary prevention) and it is known who has been prescribed lipid lowering drugs at the Ian Charleson Day Centre, I have no access to information about drugs prescribed by general practitioners. Thus the proportion receiving lipid lowering drugs is likely to be substantially underestimated.

In this chapter and in Chapter 4 I have created a definition of a TC event, and investigated the proportion that has met this definition. I could instead have considered the average change in total cholesterol at a particular time point. This type of analysis is also likely to be affected by many of the same biases described here, as patients require the presence of a measurement at a particular time point to be included in these analyses.

Although increases in total cholesterol when starting HAART were common in this cohort, on only three occasions did this lead to an individual stopping an ARV for the reason of a lipid abnormality. This implies that although changes in total cholesterol appear to be clinically significant in terms of the increase in risk of cardiovascular disease, it may be that interventions other than changing antiretrovirals are being

considered to treat the condition, at least in the short-term. Therefore, considering the correlation between treatment change and the definition of hypercholesterolaemia may not be of great importance. Nonetheless, it is perhaps a little concerning that so few of the patients discontinuing an ARV for the reason of a lipid abnormality met some of the definitions for hypercholesterolaemia proposed in Table 5.4. Therefore, a number of patients with a toxicity thought clinically important by their treating clinician were not identified by our endpoint. However, in this chapter we considered the proportion experiencing a total cholesterol event at the time of their first measurement in the window six months to one year after starting HAART. The three patients who made changes to their regimen after this time, and so perhaps they were correctly classified at the time point we were considering.

When investigating the prevalence of hepatotoxicity, by considering AST and ALT levels, much lower toxicity rates were observed than that seen for hypercholesterolaemia. Nonetheless, there was still variability in the proportion considered to be experiencing hepatotoxicity depending on the endpoint considered. The endpoints in which a confirmed high AST/ALT level were required led to considerably lower estimates of the prevalence of hepatotoxicity compared to when a single value during the first year of follow-up was required. The highest prevalence was seen when considering an AST/ALT measurement >2.5 times ULN higher than pre-HAART levels as an event. Again, we must weigh up the benefits of considering an endpoint in which all those with evidence of hepatotoxicity are captured against one which is non-specific and thus a large proportion of individuals who are not truly experiencing hepatotoxic toxicity are considered to be doing so.

As the proportion of patients experiencing a hepatotoxicity event was so low, the analyses that may be performed comparing differences between different demographic subgroups may not be possible. One potential solution is to use a different endpoint, such as any high measurement in the first year of HAART. Although I have observed in Chapter 4 that this is likely to be associated with a greater bias than that seen when using the first measurement in a window, as used in this chapter, it may be that using an endpoint in which a greater number of events is observed may still be preferable. Furthermore, I identified in Chapter 4 that AST/ALT monitoring was not as strongly affected by calendar time as was total cholesterol levels. As the next 'best' endpoint identified in Chapter 4 was any high measurement in the first year of HAART, it may be preferable to use this endpoint instead to increase the power of the study, whilst bearing in mind the potential biases related to frequency of monitoring.

In the hepatotoxicity setting most, although still not all, patients had a pre-HAART AST and ALT measurement, as well as at least one follow-up measurement, and therefore the issue of biases relating to who is included is likely to have less impact than when considering cardiovascular events. However, around 30% of patients could not be included when a pre-HAART value and at least one follow-up measurement were required. This number was not greatly reduced when two measurements in the first year of HAART were required. This suggests that the requirement of a confirmed endpoint is not unreasonable. However, when considering those who discontinued an antiretroviral in the first year of HAART for the reason of abnormal LFTs, it is noticeable that the median (range) AST and ALT values were 174 (34, 491) IU/L and 194 (30, 953) IU/L respectively. As 5 times the ULN corresponds to a value of 200 IU/L (the ULN is 40 IU/L for both markers), this means that more than 50% of those discontinuing an ARV did so before reaching this level. Thus, it is likely that clinicians are intervening before AST/ALT reach these cut offs and are considering AST and ALT values at this level to be clinically significant. They may also be likely to intervene before a second high AST/ALT measurement is recorded, meaning that a number of true hepatotoxicity events would be missed if a confirmed endpoint were required. This is reflected in Table 5.9 which showed that few of the patients stopping an ARV for abnormal LFTs were considered to have an event regardless of the endpoint chosen. Furthermore, the median (range) AST and ALT levels pre-HAART were 39 (24, 159) IU/L and 52 (18, 167) IU/L respectively, indicating that more than half had high levels pre-HAART. This suggests that inclusion of all individuals, even those with high pre-HAART levels, is necessary to capture all those at risk of developing hepatotoxicity. Thus, considering the change from pre-HAART levels for a single measurement, such as in definition 5 in Table 5.7 may be more appropriate for this endpoint.

5.8 Summary

In conclusion, in this chapter I have shown that the inclusion of dose reductions and formulation changes when estimating the proportion discontinuing at least one ARV in the first year of HAART does not substantially alter the results obtained. The inclusion of these events may be a benefit if they reflect reductions due to toxicity reasons, but if they merely reflect changes for convenience then the proportion discontinuing ARVs may be over estimated. When considering laboratory-defined adverse events, the cut-offs and definitions used have a great impact in the prevalence observed, and this may explain to some extent the disparate results observed in Chapter 2. Therefore, the clinical implications of the endpoint chosen and the specific situation being investigated must be considered carefully prior to any study being carried out.

Chapter 6 – The relationship between CD4 cell count nadirs and the toxicity profiles of antiretroviral regimens

6.1 Introduction

In previous chapters I have investigated the potential biases that may be present when investigating the prevalence and incidence of antiretroviral-related toxicities, and identified the endpoints that may be least affected by these biases. A further issue identified in Chapter 2 that may influence the observed prevalence of antiretroviral-related toxicities from different studies is the demographic characteristics of patients included. By assessing the impact of co-factors on the occurrence of toxicities, we will be able to investigate whether this is a likely explanation for differences in the prevalence of toxicities observed in studies carried out in different population groups. It is also important to investigate the association between potential confounders and the occurrence of toxicities, and thus to identify whether certain individuals are at a greater risk of experiencing toxicities than others.

In this chapter, I shall investigate the associations between several potential confounders and the occurrence of antiretroviral-related toxicities, focusing in particular on the pre-HAART CD4 cell nadir. There have been a number of previously published studies suggesting that a lower CD4 cell nadir (the lowest CD4 cell count measurement recorded for an individual prior to starting HAART) is associated with an increased prevalence of toxicities amongst those on HAART [119;375-379]. However, others have not found such an association [199;213;333;380;381]. Although the primary concern when considering when to start HAART is the virological and immunological, (and consequently clinical) response to HAART, if such a relationship does exist then this could be a further contributory factor that would potentially impact on the patient's and clinician's decision of when to start treatment as individuals may wish to start treatment at higher CD4 cell counts to minimise the risk of toxicity occurring.

This chapter will also enable me to apply the results from chapters 4 and 5, considering the most appropriate endpoints to use. I will concentrate firstly on all toxicities, by investigating discontinuations of antiretrovirals for toxicity reasons. Subsequently, I shall focus on the occurrence some of the most common laboratory-defined toxicities; hypercholesterolaemia, hypertriglyceridaemia, hepatotoxicity, hyperbilirubinaemia and anaemia.

6.2 Methods

6.2.1 Patient population

As in previous chapters, patients included in this study were from the Royal Free Hospital. Patients were previously antiretroviral-naïve and starting HAART between 1st January 1998 and 1st January 2005. Further, individuals were required to have a CD4 cell count and viral load measured in the period six months before starting HAART. Thus, the patients included are identical to those described in Chapter 5. The total cholesterol measurement taken in the six-month period before starting HAART that was measured closest to the date of starting HAART was taken as the baseline total cholesterol measurement. The baseline triglyceride, AST, ALT, haemoglobin, bilirubin, CD4 cell count and viral load measurements were defined similarly.

6.2.2 Treatment-limiting adverse events

I firstly considered the occurrence of toxicities that led to the discontinuation of an antiretroviral drug. I divided the reasons for stopping each antiretroviral, as recorded from the patient notes, into categories identical to those used in Chapter 5. Patients were considered to have discontinued a drug for toxicity reasons if any of the following were listed as a reason for discontinuation: renal problem, nausea, anaemia, mouth ulcer, lipodystrophy, abdominal pain, malaise, gastrointestinal (GI) side effects, peripheral neuropathy, central nervous system (CNS) disorders, depression, abnormal liver function test (LFT) results, intolerance, rash, allergy, diarrhoea, diabetes/raised glucose, lipid abnormality, pancreatitis, headache, myositis, raised amylases, lactic acidosis or skin problems. Patients were considered to have ceased an antiretroviral for efficacy reasons if the reason given for stopping an antiretroviral was "failure-viral load", "failure-CD4" or "failure-resistance". All other reasons for stopping were classed as "other" and consisted of "only receiving the drug to prevent mother to child transmission", tetragenic effects during pregnancy, "drug interaction", "rationalisation", "poor compliance", "patient choice", "study end point", "death", "not recorded", or "other". The term "rationalization" is used in our clinic for switches such as AZT+3TC to combivir, or discontinuation of one drug when changes are made to other drugs in the regimen resulting in the unsuitability of the drug.

I considered the date of first discontinuation or switch of any antiretroviral in the HAART regimen. Dose and formulation changes were included in these analyses, in case they were toxicity-related (as investigated in Chapter 5, this assumption is unlikely to have a major impact on the study results). Clinicians are able to list up to three reasons for discontinuing each antiretroviral, and some individuals discontinued more

than one antiretroviral on the same date, for which they were able to give different reasons for discontinuing different antiretrovirals. However, for this analysis, I wished to consider only a single reason for drug discontinuation or switch on the date that the individual first made a change to their antiretroviral regimen. In 183 cases of the first date of stopping or switching an antiretroviral, more than one reason for stopping was given. Therefore, the following rules were applied to choose a single reason for drug discontinuation/ switch:

(1) If two reasons for drug change/discontinuation were given, one of the reasons fell into the "other" category (as described in the previous paragraph) and the second reason fell into the "efficacy" or "toxicity" category for stopping, then the "efficacy" or "toxicity" reason was chosen.
(2) If two reasons for drug change/discontinuation were given, one of the reasons fell into the "efficacy" category and the second reason fell into the "toxicity" category for stopping, then the "toxicity" reason was chosen.
(3) If two reasons for drug change/discontinuation were given, one reason given was patient choice and the other reason was rationalization, then patient choice was chosen.
(4) If two reasons for drug change/discontinuation were given, one reason given was rationalization and the other reason given was 'other', then rationalization was chosen.
(5) If two reasons for drug change/discontinuation were given, one reason given was patient choice and the other reason given was 'other', then patient choice was chosen.
(6) If two reasons for drug change/discontinuation were given, one reason given was poor compliance and the other reason was other, then poor compliance was chosen as the reason for stopping

After applying thee general rules, there were 36 occasions where more than one toxicity reason was given on the same date for drug discontinuation. The reasons that occurred, and the final reason for first drug discontinuation/switch chosen are shown in Table 6.1. Although the choice is subjective, I have attempted to choose those toxicities that are associated with more serious long-term consequences.

Table 6.1 – Reasons chosen for first drug discontinuation/switch when more than one "toxicity" reason was given for stopping/switching an antiretroviral drug on the same date

Reasons for antiretroviral discontinuation given			Reason chosen
1	2	3	
CNS effects/insomnia	Nausea/vomiting		CNS effects/ insomnia
CNS effects/insomnia	Lipodystrophy		CNS effects/ insomnia
CNS effects/insomnia	GI side effects		CNS effects/ insomnia
CNS effects/insomnia	Headache	Nausea/vomiting	CNS effects/ insomnia
Lipid abnormality	Lipodystrophy	Myositis	Lipid abnormality
Lipid abnormality	Lipodystrophy		Lipid abnormality
Lipid abnormality	Diarrhoea		Lipid abnormality
Peripheral neuropathy	Lipodystrophy		Peripheral neuropathy
Peripheral neuropathy	Diarrhoea		Peripheral neuropathy
Peripheral neuropathy	Lipid abnormality		Peripheral neuropathy
Abnormal LFTs	Rash		Abnormal LFTs
Abnormal LFTs	Nausea/vomiting	Rash	Abnormal LFTs
Abnormal LFTs	Lipodystrophy		Abnormal LFTs
Abnormal LFTs	Lipid abnormality		Abnormal LFTs
Abdominal pain	GI side effects		Abdominal pain
Abdominal pain	Nausea/vomiting		Abdominal pain
Myositis	Malaise/fatigue		Myositis
Diabetes/raised glucose	Diarrhoea		Diabetes/raised glucose
Nausea/vomiting	Diarrhoea		Nausea/vomiting
Headache	Nausea/vomiting		Headache
Malaise/fatigue	Nausea/vomiting		Malaise/fatigue
Skin problems	Anaemia	Nausea/vomiting	Skin problems
Anaemia	Malaise/fatigue		Anaemia
Diarrhoea	Nausea/vomiting		Diarrhoea
Rash	Diarrhoea		Rash
GI side effects	Diarrhoea		GI side effects
Allergic reaction	Rash		Allergic reaction

Ordering of reasons (i.e. 1, 2, 3) does not necessarily reflect their importance

I then used standard survival methods to investigate the relationship between the CD4 cell nadir and the time to a treatment-limiting adverse event, as defined by the first drug discontinuation or switch in the HAART regimen. Other potential explanatory factors considered were: age (per 10 year increment), ethnicity (white, black African and other), risk group for HIV transmission (homosexual, heterosexual and other), HAART regimen (1PI+2NRTI, 1NNRTI+2NRTI, 1PI+RTV+2NRTI and other), baseline viral load (per 1 log copies/ml higher), gender and calendar year of starting HAART (per year later). Individuals who have never discontinued an antiretroviral in their regimen were censored at their last date of follow-up, which was defined as the date of the last recorded CD4 cell measurement for that individual. As I wished to investigate the factors associated with stopping treatment for toxicity reasons, those who first discontinued an antiretroviral for efficacy or other reasons were dealt with in three

different ways. The first analysis considered the first discontinuation of an antiretroviral for any reason as an event, which is comparable to the standard "intention to treat" analysis. The second censored those whose first discontinuation of an antiretroviral in their HAART regimen was for efficacy or other reasons at the time of discontinuation and is comparable to an "on treatment" analysis. The third method censored those whose first discontinuation of an antiretroviral in their HAART regimen was for efficacy or other reasons at the end of their follow-up period. This method has been developed to account for the presence of competing risks [370], which were briefly mentioned in Chapter 5. Competing risks are said to occur when the occurrence of one event means that the event of interest can no longer occur. Here, stopping an antiretroviral for efficacy reasons or for other reasons means that an individual is no longer able to stop the same antiretroviral for toxicity reasons, and thus is no longer at risk of experiencing the event of primary interest (stopping a drug for toxicity). If an individual is censored at this time point, then the analysis is likely to underestimate the chances of experiencing an event, and censoring individuals who experience a competing event at the end of the follow-up period, rather than at the time of the competing event, accounts for this. However, the interpretation of the relative hazards obtained from a proportional hazards model is not so straight forward in a competing risk model. The relative hazards become cause specific hazards, and this models the occurrence of toxicity-related antiretroviral discontinuations in the absence of the occurrence of efficacy- and other-related antiretroviral discontinuations [382-384]. I then compared the results of these three methods to investigate the impact of the analytical approach on the results obtained.

6.2.3 Laboratory-defined adverse events

I then investigated the occurrence of laboratory-defined adverse events. I chose to investigate the following laboratory markers: AST and ALT levels, haemoglobin, total cholesterol, anaemia, bilirubin and triglycerides. As individuals were required to have either a baseline total cholesterol, AST/ALT, haemoglobin, bilirubin or triglyceride measurement prior to starting HAART to be included in each analysis and at least one follow-up measurement in the period six months to one year after starting HAART, the numbers of patients included in these analyses varies according to the adverse event being investigated. Therefore, I began by comparing the characteristics of those included in each analysis, compared to the total study population.

As some individuals may have started HAART with levels of some laboratory markers outside of the normal range, I initially focused on changes from baseline values. This is

also in agreement with Chapter 5 which suggested that, with the hepatotoxicity endpoints in particular, changes in laboratory markers may be a more appropriate endpoint. Increases and decreases that were considered to be clinically significant and used in the analyses are shown in Table 6.2. To investigate any potential biases due to frequency of monitoring, I considered the number of measurements of each laboratory marker in the first year of HAART stratified by the CD4 cell nadir. I investigated the association between the CD4 cell nadir and the occurrence of a laboratory-defined toxicity at the time of the first laboratory marker in the period six months to one year after starting HAART, as suggested by Chapter 4. I used logistic regression to investigate this, both before and after adjusting for the potential confounders listed in Subsection 6.2.2.

As a sensitivity analysis, I then modified the definition of the laboratory-defined adverse events to ensure that levels of the laboratory marker were above traditional cut-offs associated with a high risk of an adverse event. I then repeated the analyses described above to investigate the impact of these changes. The definitions of an adverse event are described in the right hand column of Table 6.2.

Table 6.2 – Definitions used for laboratory-defined adverse events

Event	Marker used	Definition of adverse event	
		Main analysis	Sensitivity analysis
Hypercholesterolaemia	Total cholesterol	>1 mmol/l greater than the baseline value	>1 mmol/l greater than the baseline value and >6.2 mmol/l
Hypertriglyceridaemia	Triglycerides	>1 mmol/l greater than the baseline value	>1 mmol/l greater than the baseline value and >2.3 mmol/l
Hepatotoxicity	AST and/or ALT	>2.5 times the ULN (=40 IU/L) greater than the baseline value	>2.5 times the ULN (=40 IU/L) greater than the baseline value and >5 times the ULN
Hyperbilirubinaemia	Bilirubin	>2.5 times the ULN (=17 Umol/l) greater than the baseline value	>2.5 times the ULN (=40 IU/L) greater than the baseline value and >5 times the ULN
Anaemia	Haemoglobin	>2 g/dl lower than the baseline value	Women: >2 g/dl lower than the baseline value and <11.5 g/dl Men: >2 g/dl lower than the baseline value and <13.5 g/dl

AST=aspartate aminotransferase; ALT=alanine aminotransferase; ULN=upper limit of normal

6.3 Results

6.3.1 Patient population

As in Chapter 5, 978 individuals who started first-line HAART regimens at the Royal Free Hospital were eligible for inclusion. The characteristics of these individuals at the time of starting HAART have already been described in Table 5.1. The median (IQR) pre-HAART CD4 cell nadir was 173 (71, 261) cells/mm^3. The minimum and maximum CD4 nadir observed were 0 and 1194 cells/mm^3 respectively. The median difference between the CD4 cell nadir and the baseline CD4 cell count was 0 (IQR 0, 19) cells/mm^3. In total, 68 (7.0%) individuals had a baseline CD4 cell count that was more than 100 cells/mm^3 greater than their CD4 cell nadir, and 20 (2.0%) had a baseline CD4 cell count that was more than 200 cells/mm^3 greater than their CD4 cell nadir.

6.3.2 Treatment-limiting adverse events

In total 464 (47.4%) individuals stopped at least one antiretroviral within 48 weeks of starting HAART. Table 6.3 shows the reasons given by the clinician for first discontinuation of an antiretroviral within the first 48 weeks of HAART, stratified by CD4 cell nadir (with multiple reasons for stopping classified as described in section 6.2.1). The reasons for stopping were varied. Few individuals (30; 3.1%) cited viral load or CD4 cell count failure as the reason for stopping an antiretroviral within the first 48 weeks of HAART. Fourteen (4.3%) of 323 individuals who started HAART with a CD4 nadir of between 0 and 100 cells/mm^3, compared to 8/247 (3.2%) and 8/408 (2.0%) with a CD4 nadir of 100-200 and more than 200 cells/mm^3, respectively, discontinued an antiretroviral within 48 weeks of starting HAART giving the reason for stopping as failure. The numbers stopping an antiretroviral for toxicity reasons in each of the CD4 nadir groups were 73 (22.6%), 58 (23.5%) and 84 (20.6%) and the numbers stopping for other reasons were 69 (21.4%), 47 (19.0%) and 103 (25.3%) respectively. The exact reasons given for discontinuing an antiretroviral for toxicity reasons were varied. There were few differences between CD4 cell nadir groups with respect to the main (i.e. toxicity, efficacy or other) reason for stopping (p=0.31; chi squared test). However, the numbers of individuals who stopped their first antiretroviral as they had only started the drug to prevent mother to child transmission was, perhaps not surprisingly, only cited as a reason for drug discontinuation amongst those with a CD4 cell nadir of 200+ cells/mm^3 (in 12 [2.9%] cases).

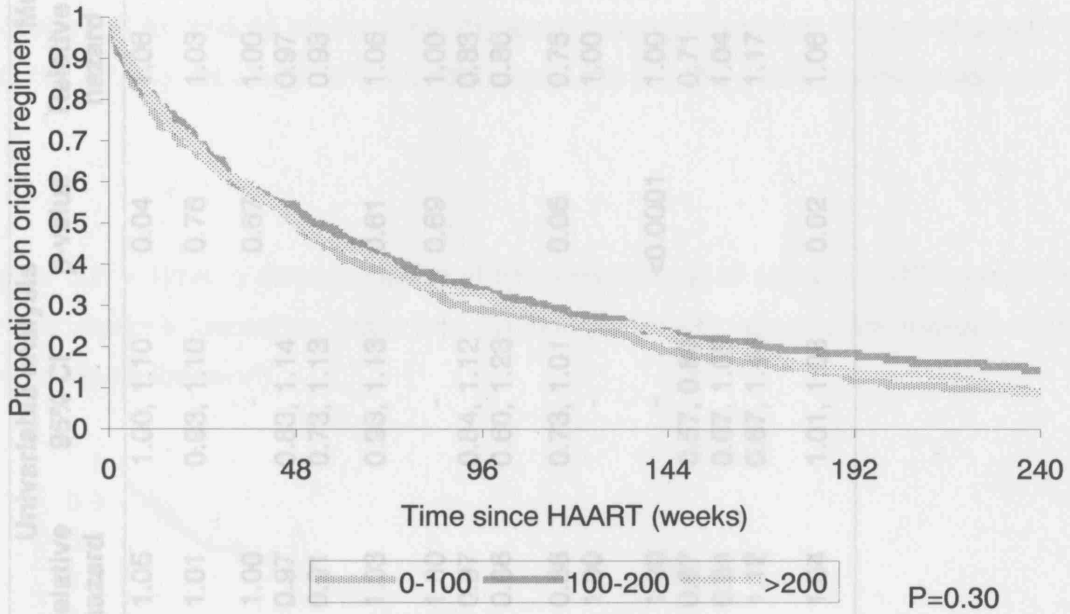
Table 6.3 – Reasons for stopping the first antiretroviral within 48 weeks of starting first-line HAART regimen

	CD4 cell nadir (cells/mm^3)			
	0-100	**100-200**	**>200**	**Total**
N	**323 (100.0)**	**247 (100.0)**	**408 (100.0)**	**978 (100.0)**
Efficacy reasons:	*14 (4.3)*	*8 (3.2)*	*8 (2.0)*	*30 (3.1)*
Toxicity reasons:				
Nausea	9 (2.8)	7 (2.8)	15 (3.7)	31 (3.2)
Anaemia	16 (5.0)	6 (2.4)	18 (4.4)	40 (4.1)
Peripheral neuropathy	15 (4.6)	7 (2.8)	3 (0.7)	25 (2.6)
CNS disorder	7 (2.2)	17 (6.9)	18 (4.4)	42 (4.3)
Rash	8 (2.5)	9 (3.6)	7 (1.7)	24 (2.5)
Diarrhoea	3 (0.9)	2 (0.8)	6 (1.5)	11 (1.1)
Abnormal LFTs	5 (1.6)	1 (0.4)	2 (0.5)	8 (0.8)
Allergy	0 (0.0)	4 (1.6)	4 (1.0)	8 (0.8)
Malaise	2 (0.6)	1 (0.4)	1 (0.3)	4 (0.4)
Headache	1 (0.3)	0 (0.0)	1 (0.3)	2 (0.2)
Renal problem	2 (0.6)	0 (0.0)	2 (0.5)	4 (0.4)
Pancreatitis	1 (0.3)	0 (0.0)	1 (0.3)	2 (0.2)
Lipodystrophy	1 (0.3)	1 (0.4)	2 (0.5)	4 (0.4)
Abdominal pain	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.1)
GI side effects	0 (0.0)	1 (0.4)	2 (0.5)	3 (0.3)
Lipid abnormality	0 (0.0)	0 (0.0)	2 (0.5)	2 (0.2)
Lactic acidosis	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.1)
Skin problems	1 (0.3)	2 (0.8)	0 (0.0)	3 (0.3)
Total	*73 (22.6)*	*58 (23.5)*	*84 (20.6)*	*215 (22.0)*
Other reasons:				
Only to prevent vertical transmission	0 (0.0)	0 (0.0)	12 (2.9)	12 (1.2)
Patient choice	17 (5.3)	14 (5.7)	39 (9.6)	70 (7.2)
Rationalisation	16 (5.0)	9 (3.6)	20 (4.9)	45 (4.6)
Other	10 (3.1)	10 (4.1)	13 (3.2)	33 (3.4)
Not recorded	11 (3.4)	4 (1.6)	10 (2.5)	25 (2.6)
Study end point	6 (1.9)	3 (1.2)	7 (1.7)	16 (1.6)
Poor compliance	5 (1.6)	4 (1.6)	2 (0.5)	11 (1.1)
Drug interaction	3 (0.9)	3 (1.2)	0 (0.0)	6 (0.6)
Following a TDM result	1 (0.3)	0 (0.0)	0 (0.0)	1 (0.1)
Total	*69 (21.4)*	*47 (19.0)*	*103 (25.3)*	*219 (22.4)*
Total discontinuations	**156 (48.3)**	**113 (45.7)**	**195 (47.8)**	**464 (47.4)**

I next considered the proportion of individuals discontinuing an antiretroviral for any reason according to the pre-HAART CD4 nadir. Figure 6.1 shows a Kaplan-Meier plot of the time to discontinuation for any reason of the first antiretroviral in the initial HAART regimen. There were no significant differences according to CD4 nadir as to the time to discontinuation (p=0.30; log rank test). After 48 weeks of HAART, 49.8%

(95% CI 44.3%, 55.4%) with a CD4 cell nadir <100 cells/mm^3 had discontinued at least one antiretroviral; this proportion increased to 71.2% (66.1%, 76.3%) by 96 weeks. For those with a CD4 cell nadir of 100-200 cells/m^3 these proportions were 47.1% (40.7%, 53.4%) and 66.2% (60.0%, 72.3%) respectively; for those with a CD4 cell nadir >200 cells/mm^3 the percentages were 49.2% (44.2%, 54.2%) and 67.1% (62.4%, 71.8%).

Figure 6.1 – Time to discontinuation of first antiretroviral in initial HAART regimen for any reason



N						
<100	323	153	77	41	20	11
100-200	247	123	71	38	25	14
>200	408	194	113	64	33	17

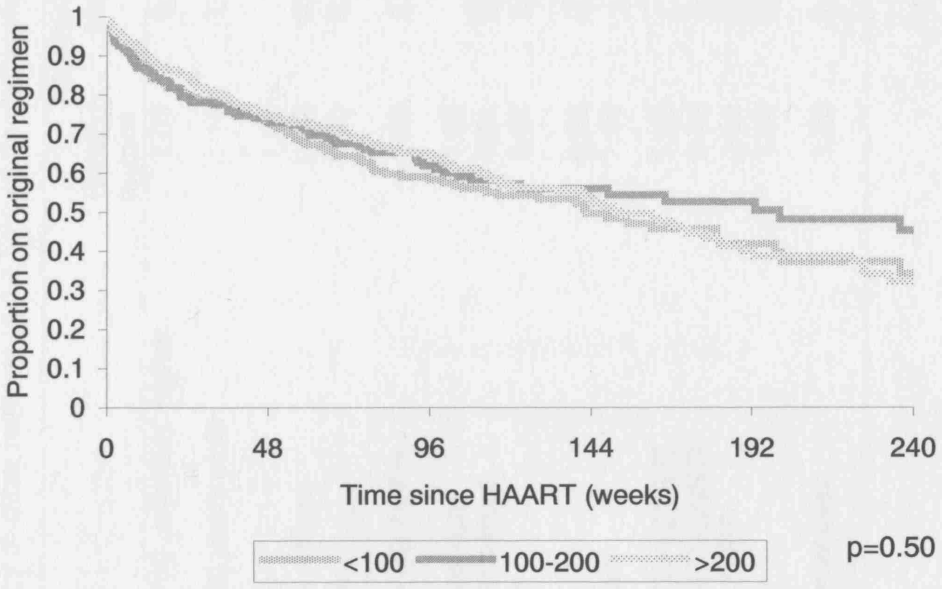
Table 6.4 shows factors associated with discontinuation of the first antiretroviral in the starting HAART regimen. In an unadjusted analysis, for each 100 cell increment in the nadir CD4 cell count, the hazard of discontinuing the first antiretroviral increased by 5% (hazard ratio [HR]=1.05; 95% CI 1.00, 1.10; p=0.04). After adjustment for potential confounders, the hazard ratio became 1.06 (1.01, 1.11; p=0.02), indicating that the hazard of stopping or switching the first antiretroviral for any reason increased slightly by 6% for each 100 cell increment in the CD4 cell nadir. Other factors found to be associated with treatment discontinuation/switch were gender, with males being 25% less likely to discontinue/switch an antiretroviral than women (HR=0.75; 0.60, 0.94; p=0.01), later calendar year of starting HAART (HR=1.06 per year later; 1.01, 1.11; p=0.01) and HAART regimen, with those receiving 1NNRTI+2NRTI being 29% less likely to discontinue/switch an antiretroviral than those receiving 1PI+RTV+2NRTI (HR=0.71; 0.60, 0.85).

Table 6.4 – Factors associated with time to discontinuation of first antiretroviral in initial HAART regimen for any reason

		Univariable analysis			Multivariable analysis		
		Relative hazard	**95% CI**	**p-value**	**Relative hazard**	**95% CI**	**p-value**
CD4 cell nadir	*Per 100 cells/mm^3 higher*	1.05	1.00, 1.10	0.04	1.06	1.01, 1.11	0.02
Age	*Per 10 years older*	1.01	0.93, 1.10	0.76	1.03	0.94, 1.12	0.54
Ethnicity	*White*	1.00	-	0.67	1.00	-	
	Black African	0.97	0.83, 1.14		0.97	0.76, 1.23	
	Other	0.91	0.73, 1.13		0.93	0.74, 1.18	
Pre-HAART viral load	*Per 1 log higher*	1.03	0.93, 1.13	0.61	1.06	0.96, 1.17	0.25
Risk group	*Homosexual*	1.00	-	0.69	1.00	-	
	Heterosexual	0.97	0.84, 1.12		0.83	0.65, 1.07	
	Other	0.86	0.60, 1.23		0.86	0.59, 1.23	
Sex	*Male*	0.86	0.73, 1.01	0.06	0.75	0.60, 0.94	0.01
	Female	1.00	-		1.00	-	
Type of HAART regimen	*1PI+RTV+2NRTI*	1.00	-	<0.0001	1.00	-	
	1NNRTI+2NRTI	0.67	0.57, 0.80		0.71	0.60, 0.85	
	1PI+2NRTI	0.84	0.67, 1.05		1.04	0.79, 1.36	
	Other	1.12	0.87, 1.44		1.17	0.90, 1.52	
Year of starting HAART	*Per year later*	1.04	1.01, 1.08	0.02	1.06	1.01, 1.11	0.01

I next considered the time to first drug discontinuation or switch, but this time censored those who discontinued a drug for efficacy or other reasons at the time of drug discontinuation, to focus more closely on those who were experiencing drug toxicity. A Kaplan-Meier plot showing the results of this analysis is shown in Figure 6.2. There was no evidence from this analysis that there were different rates of drug discontinuation according to the CD4 cell nadir (p=0.50; log rank test). After 48 weeks of HAART 26.7% (95% CI 21.3%, 32.0%) of those with a CD4 cell nadir <100 cells/mm^3, 26.5% (20.5%, 32.4%) of those with a nadir of 100-200 cells/mm^3 and 25.4% (20.6%, 30.2%) of those with a nadir >200 cells/mm^3 had discontinued an antiretroviral for toxicity reasons; by 96 weeks these percentages increased to 41.7% (35.1%, 48.3%), 38.1% (30.8%, 45.4%) and 34.8% (29.1%, 40.5%) respectively.

Figure 6.2 – Time to discontinuation of first antiretroviral in initial HAART regimen for toxicity reasons (censoring those who discontinue for efficacy or other reasons at the time of discontinuation)



N						
<100	323	153	77	41	20	11
100-200	247	123	71	38	25	14
>200	408	194	113	64	33	17

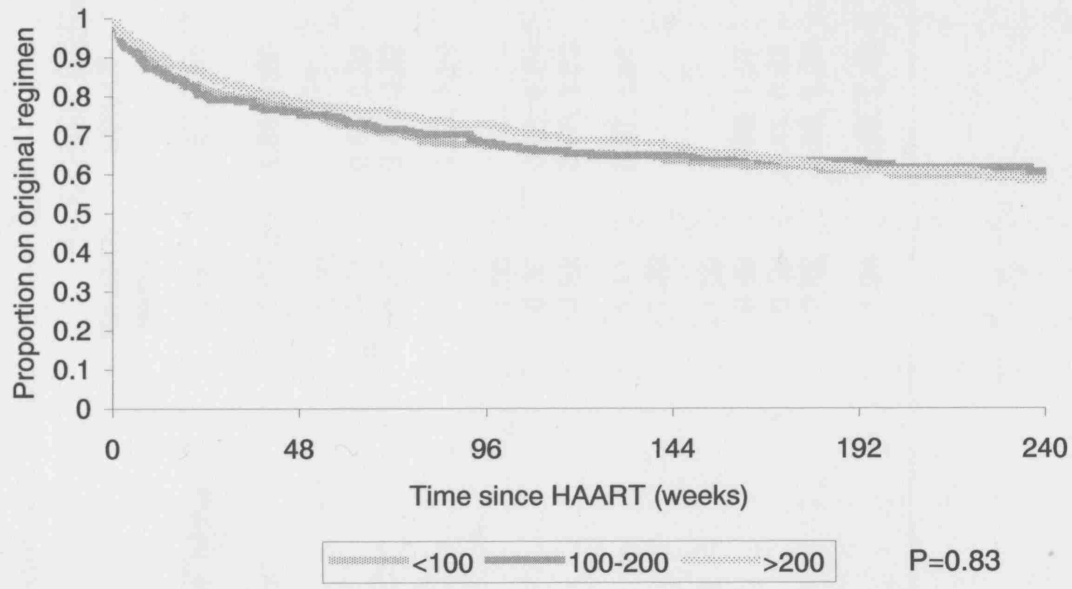
When carrying out a Cox proportional hazards regression model to investigate factors associated with the first antiretroviral discontinuation for toxicity reasons, censoring those who discontinued for efficacy or other reasons at the time of discontinuation (Table 6.5), I found no association between the CD4 cell nadir and time to discontinuation, either in univariable analysis (HR=1.03; 0.96, 1.10; p=0.41) or in multivariable analysis (HR=1.02; 0.95, 1.09; p=0.62). Older age was associated with an increased hazard of a first treatment discontinuation/switch (HR=1.21 per 10 years older; 1.07, 1.37; p=0.002). Those of other ethnicities (HR=0.63 compared to white; 0.42, 0.93), with a heterosexual risk for HIV transmission (HR=0.66 compared to homosexual; 0.45, 0.99) and of female gender (HR=0.63 for males compared to females; 0.44, 0.90; p=0.01) had a decreased hazard of drug discontinuation/switch.

Table 6.5 –Factors associated with time to discontinuation of first antiretroviral in initial HAART regimen for toxicity reasons (censoring those whose first discontinuation was for efficacy or other reasons at the time of discontinuation)

		Univariable analysis			**Multivariable analysis**		
		Relative hazard	95% CI	**p-value**	**Relative hazard**	**95% CI**	**p-value**
CD4 cell nadir	*Per 100 cells/mm^3 higher*	1.03	0.96, 1.10	0.41	1.02	0.95, 1.09	0.62
Age	*Per 10 years older*	1.18	1.05, 1.32	0.005	1.21	1.07, 1.37	0.002
Ethnicity	*White*	1.00	-	0.01	1.00	-	
	Black African	0.83	0.65, 1.05		0.92	0.64, 1.32	
	Other	0.60	0.42, 0.87		0.63	0.42, 0.93	
Pre-HAART viral load	*Per 1 log higher*	0.98	0.85, 1.13	0.79	0.98	0.85, 1.14	0.80
Risk group	*Homosexual*	1.00	-	0.15	1.00	-	
	Heterosexual	0.81	0.65, 1.01		0.66	0.45, 0.99	
	Other	0.88	0.53, 1.46		0.83	0.50, 1.40	
Sex	*Male*	0.96	0.75, 1.23	0.75	0.63	0.44, 0.90	0.01
	Female	1.00			1.00	-	
Type of HAART regimen	*PI+RTV+2NRTI*	1.00	-	0.05	1.00	-	
	1NNRTI+2NRTI	0.78	0.61, 1.00		0.83	0.64, 1.07	
	1PI+2NRTI	0.63	0.43, 0.91		0.76	0.49, 1.17	
	Other	1.00	0.67, 1.49		1.00	0.66, 1.50	
Year of starting HAART	*Per year later*	1.06	1.00, 1.12	0.05	1.04	0.97, 1.11	0.29

Finally, when taking a competing risks approach to account for those discontinuing an antiretroviral for efficacy or other reasons, by censoring their follow-up at the end of the treatment period, I again found no association between the CD4 nadir and discontinuing an antiretroviral for toxicity reasons (Figure 6.3; p=0.83 log-rank test). After 48 weeks of treatment 23.4% (18.7%, 28.1%), 24.2% (18.8%, 29.7%) and 21.6% (17.5%, 25.7%) with a CD4 cell nadir of <100 cells/mm^3, 100-200 cells/mm^3 and >200 cells/mm^3, respectively, had discontinued an antiretroviral for toxicity reasons; at 96 weeks these percentages increased to 32.6% (27.4%, 37.9%), 31.8% (25.8%, 38.4%) and 27.1% (22.6%, 31.5%).

Figure 6.3 – Time to discontinuation of first antiretroviral in initial HAART regimen for toxicity reasons (censoring those who discontinue for efficacy or other reasons at the end of follow-up)



N						
<100	323	213	178	138	101	68
100-200	247	170	131	98	77	52
>200	408	296	246	176	123	88

In a Cox proportional hazards model (Table 6.6), there was no evidence that the CD4 cell nadir was associated with drug discontinuation for toxicity reasons, either in a univariable (HR=1.00; 0.94, 1.07; p=0.90) or multivariable analysis (HR=0.98; 0.91, 1.05; p=0.57). For every ten year increment in age, the cause specific hazard of experiencing treatment switch/discontinuation increased by 24% (HR=1.24; 1.09, 1.40) and males were 32% less likely to switch or discontinue (HR=0.68; 0.48, 0.97). Other

factors associated with a decreased risk of treatment discontinuation/switch were other ethnicities (HR=0.61 compared to white; 0.41, 0.90), heterosexual risk (HR=0.67; 0.45, 0.99) and use of 1PI+2NRTI regimens (HR=0.61; 0.40, 0.94).

As a sensitivity analysis, I repeated all analyses, considering the CD4 nadir on a $\log_2$ scale, rather than on a linear scale. Here, I found no evidence of an association between the CD4 nadir and treatment-limiting adverse events. In multivariable analyses I found a hazard ratio of 0.98 per 1 $\log_2$ CD4 higher (95% CI 0.94, 1.02; p=0.36) for treatment discontinuation when considering all reasons for treatment discontinuation, a hazard ratio of 0.96 (0.90, 1.03; p=0.25) when censoring those discontinuing for efficacy or other reasons at the date of discontinuation, and a hazard ratio of 0.98 (0.92, 1.04; p=0.42) when censoring those discontinuing for efficacy or other reasons at the end of follow-up.

Table 6.6 – Factors associated with time to discontinuation of first antiretroviral in initial HAART regimen for toxicity reasons (censoring those whose first discontinuation was for efficacy or other reasons at the end of follow-up)

		Univariable analysis			Multivariable analysis		
		Relative hazard	**95% CI**	**p-value**	**Relative hazard**	**95% CI**	**p-value**
CD4 cell nadir	*Per 100 cells/mm^3 higher*	1.00	0.94, 1.07	0.90	0.98	0.91, 1.05	0.57
Age	*Per 10 years older*	1.21	1.08, 1.35	0.001	1.24	1.09, 1.40	0.0007
Ethnicity	*White*	1.00	-	0.002	1.00	-	
	Black African	0.76	0.59, 0.96		0.86	0.60, 1.23	
	Other	0.57	0.40, 0.82		0.61	0.41, 0.90	
Pre-HAART viral load	*Per 1 log higher*	0.96	0.84, 1.09	0.50	0.92	0.80, 1.07	0.28
Risk group	*Homosexual*	1.00	-	0.04	1.00	-	
	Heterosexual	0.75	0.60, 0.94		0.67	0.45, 0.99	
	Other	0.92	0.55, 1.53		0.85	0.50, 1.42	
Sex	*Male*	1.11	0.87, 1.41	0.41	0.68	0.48, 0.97	0.03
	Female	1.00	-		1.00	-	
Type of HAART regimen	*1PI+RTV+2NRTI*	1.00	-	0.02	1.00	-	
	1NNRTI+2NRTI	0.99	0.78, 1.27		1.01	0.78, 1.30	
	1PI+2NRTI	0.59	0.41, 0.86		0.61	0.40, 0.94	
	Other	0.82	0.55, 1.23		0.77	0.51, 1.17	
Year of starting HAART	*Per year later*	1.04	0.99, 1.10	0.14	1.00	0.94, 1.07	0.97

6.3.3 Laboratory-defined adverse events

I next went on to consider laboratory-defined toxicities. As in chapter 6, there were 480 participants with a pre-HAART total cholesterol measurement and at least one follow-up measurement in the period six months to one year after starting HAART who could therefore be included in these analyses. These patients have already been described in Table 5.1. The median (IQR; range) CD4 cell nadir in this group was 180 (77, 261; 0, 1194) cells/mm^3. The same number of patients were eligible for the hypertriglyceridaemia analysis, with a median (IQR) pre-HAART triglyceride level of 1.4 (1.0, 2.0) mmol/l. There were 670 patients eligible for both the hepatotoxicity and hyperbilirubinaemia analysis. These patients are also described in Table 5.1. The median (IQR; range) CD4 cell nadir in these patients was 178 (73, 260; 0, 1194) cells/mm^3. The median (IQR) pre-HAART bilirubin level was 7 (5, 9) U/L. Six hundred and fifty eight individuals were eligible for the anaemia analysis. The characteristics of this subgroup are similar to the entire study population, with a median (IQR; range) pre-HAART CD4 cell nadir count of 178 (75, 260; 0, 1194) cells/mm^3. The median (IQR) pre-HAART haemoglobin level was 13.0 (11.5, 14.3) g/dl.

There was no association between the CD4 cell nadir and the number of total cholesterol ($p=0.25$) and triglyceride ($p=0.32$) measurements taken in the first year of HAART (Table 6.7). However, those with a CD4 cell nadir <100 cells/mm^3 had, on average, more AST, ALT, bilirubin and haemoglobin measurements in the first year of HAART than those with higher nadir values ($p<0.0001$ for all markers).

Table 6.7 – Median (inter-quartile range, range) number of laboratory measurements taken in the first year of HAART, according to the CD4 cell nadir

Median IQR Range	**CD4 nadir group**			**p-value**
	<100 cells/mm^3	**100-200 cells/mm^3**	**>200 cells/mm^3**	
Total cholesterol	5 3, 7 1, 17	5 2, 6 1, 14	5 2, 6 1, 20	0.25
Triglycerides	5 3, 7 1, 17	5 2, 6 1, 14	4 2, 6 1, 20	0.32
AST	7 5, 9 1, 61	6 4, 8 1, 42	5 4, 7 1, 30	<0.0001
ALT	7 5, 9 1, 61	6 4, 8 1, 42	5 4, 7 1, 30	<0.0001
Bilirubin	7 5, 9 1, 61	6 4, 8 1, 42	5 4, 7 1, 30	<0.0001
Haemoglobin	7 5, 10 1, 80	6 4, 8 1, 61	5 4, 7 1, 32	<0.0001

p-values from Kruskal-Wallis test

The number of individuals experiencing toxicity events and the results of a logistic regression investigating factors associated with a toxicity event are shown in Table 6.8. There was evidence that a lower CD4 cell nadir was associated with an increased odds of a total cholesterol increase in multivariable analysis (OR=0.82 per 100 cells/mm^3 higher; 95% CI 0.81, 0.95; p=0.008), but no evidence of an association between the CD4 cell nadir and hypertriglyceridaemia (OR=0.99; 0.85, 1.15; p=0.90). When considering a hepatotoxicity event, there was no evidence of an association with CD4 cell nadir in univariable analysis (OR=0.81; 0.48, 1.37; p=0.43); unfortunately there were not enough events (n=9) to be able to perform an adjusted analysis. Similarly, there was no evidence of an association with hyperbilirubinaemia in unadjusted analysis (OR=0.94; 0.54, 1.64; p=0.82), but as only 6 events occurred, no multivariable analysis was performed. There was also no evidence of an association between the CD4 cell nadir and anaemia (adjusted OR=1.08; 0.85, 1.37; p=0.52).

Table 6.8 – The impact of a 100 cells/mm^3 increment CD4 cell nadir on the occurrence of antiretroviral-related toxicity within the first year of HAART, as defined by laboratory markers. Results from a logistic regression analysis

		Univariable analysis			Multivariable analysis		
	Number of events	**Odds Ratio**	**95% CI**	**p**	**Odds Ratio**	**95% CI**	**p**
Total cholesterol	173	0.77	0.68, 0.89	0.0002	0.82	0.81, 0.95	0.008
Triglycerides	88	1.07	0.93, 1.22	0.33	0.99	0.85, 1.15	0.90
AST/ALT *	9	0.81	0.48, 1.37	0.43	-	-	-
Bilirubin *	6	0.94	0.54, 1.64	0.82	-	-	-
Haemoglobin	29	1.19	0.98, 1.45	0.08	1.08	0.85, 1.37	0.52

* No multivariable model due to small number of events. Multivariable analyses adjusted for age, gender, risk group, ethnicity, pre-HAART viral load, type of HAART regimen, and calendar year. For definition of events see Table 6.2

I next carried out a sensitivity analysis considering both changes in the laboratory markers as well as the absolute value of these markers. The results from these sensitivity analyses were broadly similar to those from the primary analysis (Table 6.9). In multivariable analysis, the OR for each 100 cells/mm^3 increment in CD4 cell nadir was 0.76 for a 1 mmol/l increase in total cholesterol (95% CI 0.57, 1.00; p=0.05), and thus there was weak evidence of an association. An odds ratio of 1.01 was found for the association between the CD4 nadir and triglyceride increases (95% CI 0.86, 1.18; p=0.93). Again, the numbers experiencing a hepatic event and a bilirubin event were too small to be able to conduct multivariable analyses. Finally, I observed no association between the nadir and haemoglobin changes (adjusted OR=1.08; 0.81, 1.45; p=0.58).

Table 6.9 – Sensitivity analysis of the impact of a 100 cells/mm^3 increment in CD4 cell nadir on the occurrence of antiretroviral-related toxicity within the first year of HAART, as defined by laboratory markers.

Results from a logistic regression analysis

		Univariable analysis			Multivariable analysis		
	Number of events	**Odds Ratio**	**95% CI**	**p**	**Odds Ratio**	**95% CI**	**p**
Total cholesterol	39	0.76	0.59, 0.99	0.04	0.76	0.57, 1.00	0.05
Triglycerides	74	1.07	0.93, 1.24	0.33	1.01	0.86, 1.18	0.93
AST/ALT *	6	0.91	0.51, 1.62	0.76	-	-	-
Bilirubin *	3	0.92	0.41, 2.05	0.84	-	-	-
Haemoglobin	19	1.16	0.91, 1.48	0.24	1.08	0.81, 1.45	0.58

* No multivariable model due to small number of events. Multivariable analyses adjusted for age, gender, risk group, ethnicity, pre-HAART viral load, type of HAART regimen, and calendar year. For definition of events see Table 6.2

6.4 Discussion

These analyses have shown that the CD4 nadir appears to have little effect on the rates of first discontinuation or switching of antiretrovirals for toxicity reasons. The relative hazards obtained for the relationship between the CD4 nadir and the occurrence of treatment-limiting adverse events were close to 1 with narrow confidence intervals, implying that even if such associations exist they are likely to be of small magnitude. The exception to this was the case when considering all treatment discontinuations and switches as an event; here those with higher CD4 cell nadirs were more likely to discontinue treatment. This relationship is quantatively different to that reported in other studies, in which a higher CD4 cell nadir has been associated with an increased rate of drug switch/discontinuation. It is possible that this effect may have been driven by the "other" and "efficacy" categories, perhaps reflecting the fact that those with higher pre-HAART CD4 cell nadirs feel more comfortable to discontinue treatment. Furthermore, the treatment effect, although statistically significant, was small

in magnitude. Although there may be exceptions for some specific toxicities, there would appear to be relatively little evidence for a strong relationship between the CD4 nadir and the incidence of drug toxicity.

An advantage of considering treatment-limiting toxicities is that the Royal Free Clinic Database contains exact stop and start dates for all antiretrovirals. Therefore, these analyses should not be adversely affected by the frequency of patient monitoring, apart from the fact that those patients with lower CD4 counts may attend the clinic more frequently. Conversely, all patients were seen at a single clinic as clinicians are likely to follow similar patterns in stopping drugs due to toxicities. Furthermore, I have included dose changes and formulation changes as an event, and therefore may be overestimating the proportion of events. However, as discussed in Chapter 5, as dose changes could potentially be toxicity-related, I felt that it was important to capture these within the endpoint.

The literature investigating the impact of CD4 cell nadirs on the occurrence of toxicity is conflicting. Some studies [375-379] have found that those with lower baseline CD4 cell counts or CD4 cell nadirs are more likely to experience HAART-related toxicities, whereas others have found no such association, especially those that have considered the relationship between the baseline CD4 cell count (rather than the CD4 nadir) and toxicities [199;213;333;380;381], although these should be similar and thus this may be an unlikely explanation. These studies have generally investigated the occurrence of specific toxicities such as hypertriglycerideaemia and hepatotoxicity. I found no evidence of an association between the CD4 cell nadir and time to treatment-limiting adverse events in general. I specifically considered a number of different definitions of toxicities (by accounting for those stopping for efficacy and other reasons in three different ways) to assess whether the findings from the other studies were likely to be driven by the definitions used. My findings would suggest that this was not the case.

The occurrence of hypercholesterolaemia, was less common amongst those with higher CD4 cell nadirs, which is in agreement with one study [376]. It is possible that differences in the patient populations and patient management across studies could explain some of the differences found. Although they were severely limited due to the small number of events, there was little evidence from my analyses that liver-related toxicities, defined in this study as increases in bilirubin and hepatotoxicity, are associated with the CD4 cell nadir. Some studies have found an association between the two [294;385], whereas others have not [303;386]. I also observed no association with laboratory-defined hypertriglyceridaemia [380] and anaemia [387] and the CD4 cell nadir.

In any study reporting negative findings we must always consider whether the study had sufficient power to detect a significant result. Post-hoc power calculations are not particularly useful once a study has been carried out, as at this point the confidence intervals presented give a better indication of how precise the estimates are after the study has been completed as they incorporate the actual observed data. For the analysis investigating the association between the CD4 cell nadir and the time to discontinuation of an antiretroviral, censoring those who discontinued for efficacy or other reasons at the time of discontinuation, the hazard ratio obtained was 1.03, with a 95% confidence interval of 0.97 to 1.10. Thus, at one extreme the smallest effect size that we would reasonably expect to see is that for every one hundred cell increment in the pre-HAART CD4 cell nadir the hazard of discontinuing an antiretroviral decreases by 3%, and at the other extreme the largest effect size that we would reasonably expect to see would be a 10% increase in the hazard of discontinuing an antiretroviral for every one hundred cell increase from this study. Any hazard ratio observed in this range is unlikely to drastically change the conclusions, and so I can conclude that the effect sizes observed are reasonably precise.

My analyses investigating the association between the CD4 cell nadir and the occurrence of laboratory-defined toxicities must be viewed with caution, as in observational data such as these we cannot rule out the possibility that those with laboratory measurements taken at baseline were at a higher risk of experiencing a toxicity than those who did not have a baseline measurement, as has been investigated in previous chapters. However, I found that the sub-groups that were included in analyses were broadly similar compared to the whole study population. I also found that those with lower CD4 cell nadirs on average had more frequent monitoring of AST, ALT, bilirubin and haemoglobin levels. This may be because these tests are part of standard routine blood tests and thus are likely to be measured whenever a CD4 count is measured, whereas lipid measurements are not. More frequent monitoring is likely to occur amongst those with lower CD4 counts who are therefore at a greater risk of clinical disease. However, I chose an endpoint for these analyses that hopefully minimised these effects as much as possible as suggested by Chapter 4. Nonetheless, it is important to interpret these results with caution. We can also not rule out the possibility that those starting HAART with a high CD4 nadir are different from those with lower nadirs in ways that are not captured fully and thus confounding exists, as will be discussed in Chapter 7.

My study has, of course, further potential limitations. I have concentrated on treatment limiting adverse events, and have not investigated those that, although they have not inhibited treatment, may affect the quality of life (and possibly adherence patterns) of individuals receiving HAART. In particular, it may be of interest to look at the relationships between the occurrence of lipodystrophy and the CD4 cell nadir, as some studies have shown an association between the two [230;388]. Unfortunately, we do not routinely collect information on quality of life, lipodystrophy and non-treatment limiting adverse events, and so I was unable to investigate these issues. Furthermore, the reasons for stopping were gained from the clinic form, and so they rely on the accurate reporting of the reasons for stopping by clinicians and accurate data collection. I have also only allowed one reason for stopping to be included in analyses. It is likely that individuals often stop antiretrovirals for a variety of reasons. For example an individual may experience virological rebound due to a lack of adherence caused by the side effects of an antiretroviral. Therefore, I may be underestimating the proportion making changes to their regimens for toxicity reasons, as some of those individuals who stop an antiretroviral drug due to virological or immunological failure may also be experiencing toxicity. However, these issues should not apply to the analyses focussing on laboratory-defined toxicities presented here, which mostly confirmed the findings of the treatment-limiting adverse events analysis.

I have concentrated on five laboratory-defined toxicities, and therefore if there are any other specific toxicities that are associated with the CD4 nadir this will not be apparent from my analyses. Unfortunately, some laboratory markers have only recently begun to be routinely measured and so I do not have sufficient numbers of individuals with measurements to be able to investigate their relationship with the CD4 nadir. For example, the number of lactate measurements is small, and therefore analyses of this marker are not possible. Therefore I have concentrated on the most commonly reported adverse events. In particular, there is some evidence of a link between the CD4 nadir and lactic acidosis [379], but unfortunately not enough individuals have had routine lactate levels taken in the past to enable me to investigate this issue.

I found 80 individuals with more than 100 cells/mm^3 difference between their CD4 cell nadir and the baseline CD4 count. It is possible that some of these individuals may have started HAART at previous centres before attending the Royal Free Hospital for treatment, and this has not been reported, although information on their CD4 counts has been transferred to the Royal Free HIV cohort. However, excluding these individuals from analyses did not significantly alter the results presented here.

Examining the association between the CD4 nadir and the occurrence of antiretroviral-related toxicities has also given me an opportunity to put into practice the findings of previous chapters and attempt to minimise the potential biases that may occur when investigating antiretroviral-related toxicities. I found that it was straightforward to use methods for analysis that should hopefully capture clinically relevant toxicities and that is not influenced by different frequency of monitoring, especially as I observed that those with lower CD4 cell nadirs on average had more frequent AST, ALT, bilirubin and haemoglobin measurements. It is important to bear in mind when interpreting the results of this analysis that this could impact on the results. However, this led to limitations with regards to power, as the number of hepatoxicity and bilirbin events that occurred was extremely low, and this meant that I was unable to perform a multivariable analysis. It may be, therefore that an endpoint, such as the number of patients experiencing a toxicity event within the first year of HAART, may have to be employed. Although this endpoint was more affected by differential frequency of monitoring than the endpoint used here, it was the 'next best' endpoint, and would results in inclusion of a greater number of 'events', allowing more comparisons to be made. It is clear that there is no ideal solution here to investigating the occurrence of hepatotoxicity. However, until results from trials assessing the optimal CD4 count at which to start HAART are available, the best evidence on this issue may be from observational data such as this.

6.5 Summary

In summary, these analyses have indicated that overall there is little evidence of any association between the time to treatment-limiting adverse events and the pre-HAART CD4 cell nadir. Furthermore, those with lower CD4 cell nadirs do not appear to be at a greater risk of dyslipidaemia or anaemia, although there is some evidence that they are at a greater risk of hyperbilirubinaemia. Therefore, this issue does not appear to be a major concern for consideration when deciding when to start HAART. Furthermore, it is unlikely to explain any differences observed in the prevalence of HAART-related toxicities observed between studies.

Chapter 7 – Sample selection (Heckman) models to account for unmeasured confounders in observational databases

7.1 Introduction and aims of chapter

In Chapter 2 I investigated the potential biases that may be present both when assessing the incidence and prevalence of antiretroviral toxicities and when comparing the incidence of antiretroviral toxicities amongst those receiving different antiretroviral regimens. As mentioned in Section 2.3, although RCTs are the most appropriate setting in which to compare the occurrence of toxicity outcomes in those receiving different regimens, there are a number of reasons why this is not always possible. Therefore, it is often useful to assess antiretroviral-related toxicities in an observational setting.

In such a setting, patients are no longer allocated to a treatment at random. Thus, when we look at whether a particular treatment is associated with a more favourable outcome than another treatment, we must always consider the possibility of confounding. As described in Section 2.3, confounding occurs when there is a variable that is associated with treatment allocation and also with the outcome of interest. In this situation, we may end up with a misleading impression of the effects of treatment on outcome, and we cannot necessarily attribute any differences between groups to the treatments themselves. Statistical methods have been developed to account for confounders that are known and precisely measured and to deal with imperfectly measured ones, such as multivariable (adjusted) analyses, propensity scores and marginal structural models. Providing all potential confounders are known and have been measured, these methods give unbiased estimates of the treatment effect [389;390]. However, in practice, we can never be sure that all potential confounders have been measured (or whether it is even possible to measure all potential confounders).

There are several examples in HIV research in which there is the potential for unknown or unmeasured confounding to occur. The two most common 'third' antiretroviral drugs currently prescribed at the Royal Free Hospital are the PI lopinavir (LPV; administered with ritonavir) and the NNRTI efavirenz (EFV). Thus, it is of interest to know whether there are differences between these two drugs with respect to toxicity and tolerability. A recent study by the ICoNA study group, in which the efficacy and tolerability of LPV was compared to EFV in previously antiretroviral naïve individuals, found comparable

virological and immunological efficacy [391]. They also found that the two drugs had similar tolerability profiles when considering the time to discontinuation of LPV/EFV for toxicity reasons, with a relative hazard of 0.92 and 95% confidence interval of 0.51 to 1.64 (p=0.76). Similar results were observed when considering time to discontinuation of LPV/EFV for any reason. Consider the situation in which we wish to investigate whether receipt of an EFV-containing regimen is associated with a greater rate of early treatment discontinuation due to toxicity, compared to receipt of a LPV-containing regimen. In this example there are a number of potential confounders that should be considered. As lopinavir is known to be associated with dyslipidaemia [391-393], those with higher pre-treatment total cholesterol measurements (who are likely to be men and those of older age [363]) may be more likely to be prescribed EFV-containing regimens. However, if there is an association between older age, gender and/or total cholesterol levels and the probability of discontinuing treatment, then these factors could act as confounders. These factors are known and measured on individuals in the Royal Free cohort (although not all will have pre-treatment total cholesterol measurements – see Chapter 4) and thus can be accounted for. However, EFV has been associated with central nervous system (CNS) disorders, including nightmares and depression [369;394]. Therefore, it is possible that those individuals who are felt to be at risk of these toxicities, such as those with CNS disorders before starting treatment, may be prescribed LPV-containing regimens. If these individuals are also more likely to discontinue antiretroviral therapy, then this is a potential confounder. As information on the presence of depression and CNS disorders is not available in the Royal Free Cohort, this is likely to be an unmeasured confounder.

Sample Selection (or Heckman) models [395-397] were first developed by James Heckman to account for selection bias in the econometrics and health economics fields. For example, these methods were used to supply labour decisions, and to model the effectiveness of housing programmes and welfare experiments [397;398]. It has been proposed that sample selection models could be applied to account for unknown and unmeasured confounders in a medical setting. Shelton et al [395] have applied these models to investigate the effects of accessing dental treatment on chewing difficulties, and Carrieri et al [367] have applied these models to cohorts of HIV-positive patients when investigating the factors associated with non adherence to highly active antiretroviral therapy. They have also been applied in other medical settings [399-402] Therefore, these methods may be appropriate to account for the presence of unknown confounding when comparing antiretroviral treatments with respect to the occurrence of toxicities.

The aim of this chapter is to explore whether Sample Selection models are a useful tool for accounting for unmeasured confounding in the medical setting. Therefore, in Section 7.2 I will describe sample selection models. I will provide an intuitive description of the ideas behind these methods and their interpretation; a more formal derivation is given in Appendix B. Next, in Section 7.3, I shall examine whether these models give unbiased results in the presence of unmeasured confounding. One way to investigate this is in a simulation study, as this enables us to control the "true" level of confounding and effect of treatment on response. Therefore, I will be able to study the performance of these methods under various conditions. Finally, in Section 7.4 I shall apply sample selection methods to a real-life example to investigate how easy they are to apply and interpret in general clinical research.

7.2 Sample Selection models – an intuitive explanation

7.2.1 The situation

Consider a situation where we have an observational dataset and wish to investigate the relative effect of receiving one or other of two HAART treatment regimens (regimen A or regimen B) on a binary outcome (for example virological response or the presence of a toxicity at a particular time point). The situation described below also applies to the situation with a continuous treatment outcome. Imagine that the observational dataset contains information on the following variables:

1. Variables that are associated with patients' outcome, but are not associated with whether an individual receives regimen A or regimen B (which I will call O_1 to O_r)
2. Variables that are associated with the regimen an individual receives, but do not influence patients' outcome (which I will call T_1 to T_s)
3. Variables that are associated with both the regimen an individual receives and the patients' outcome (which I will call OT_1 to OT_t; i.e. confounders)

7.2.2 First stage model

Sample selection models consist of two stages. The first stage (which is also sometimes known as the choice model) is a standard regression model, as described in Appendix B. The dependent variable in this model is treatment allocation (i.e. receipt of regimen A as opposed to regimen B). We include as explanatory factors in this model any known predictors of treatment allocation. In the notation described in

Subsection 7.2.1, this means that the explanatory variables included in this model are variables T_1 to T_s and OT_1 to OT_t. When carrying out regression models with binary dependent variables, it is usual to perform a logistic regression analysis. This is a model with a logit link function, given by the following formula:

$$\log it(p) = \log_e\left(\frac{p}{1-p}\right)$$

However, the sample selection theory suggests that we should instead use a probit link function for the first stage model, as sample selection models rely on the use of bivariate normal distribution theory. The probit link function is given by the formula:

$$probit(p) = \Phi^{-1}(p)$$

where Φ^{-1} is the pth percentile of the standard normal distribution (see Appendix B for details).

We can then use the results of this regression model to calculate a predicted value $X\beta$, lying between -∞ and +∞, for each individual in the study. This predicted value is based on their values of the covariates included in the model (T_1 to T_s and OT_1 to OT_t) and the parameter estimates associated with each covariate. The predicted value gives us an indication of the probability that each individual received regimen A rather than regimen B. If we wish, we can back transform these predicted values using the following formulae to obtain the probability that each individual was allocated to receive regimen A:

$$P(TreatmentA) = \frac{\exp(X\beta)}{(1+\exp(X\beta))} \qquad \text{where logit link function is applied}$$

or

$$P(TreatmentA) = \Phi(X\beta) \qquad \text{where probit link function is applied}$$

Now, imagine a person with a high predicted probability of receiving regimen A based on their values of T_1 to T_s and OT_1 to OT_t, but who still received regimen B. For this individual, we are predicting their treatment choice poorly; thus it is possible that there is something about this individual that our observational dataset has not captured.

Similarly, the treatment choice for an individual who had a very low probability of receiving regimen A but still received regimen B has also been predicted poorly.

We therefore use these predicted probabilities (via the predicted values, $X\beta$) from the first stage model to calculate the **Inverse Mills Ratio**

7.2.3 Inverse Mills Ratio

The Inverse Mills ratio (IMR, λ) is given by the following formulae:

$$\lambda = \frac{\phi(-X\beta)}{\Phi(X\beta)} \qquad \text{for those receiving regimen A}$$

$$\lambda = \frac{-\phi(-X\beta)}{\Phi(-X\beta)} \qquad \text{for those receiving regimen B}$$

where $\phi(.)$ is the probability distribution density function of the standard normal distribution and $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

It can be seen from the above formulae that the IMR is a function of the predicted values obtained from the first stage model. By examining some of the properties of this variable, we can begin to understand its purpose. Consider an individual who received regimen A when their covariates T_1 to T_s and OT_1 to OT_t suggested that this was not very likely. In this situation the IMR takes a large and positive value. Conversely, if an individual did not receive regimen A when their covariates implied that they had a high probability of doing so then the IMR is large and negative. Finally, if an individual received regimen A and their predicted value suggested this was likely, or if they did not receive regimen A and their predicted value also suggested this was likely, then the IMR has a value that is close to zero.

Thus, large (in magnitude) IMR values indicate individuals whose treatment allocation is unexpected given their covariate values even after accounting for random variation. The direction of the effect (i.e. whether the IMR takes a large positive or a large

negative value) indicates whether the patient appeared like likely to receive regimen A when this was not the case (here a large, positive value is obtained), or the patient appeared likely to regimen B when in fact the patient received regimen A (here a large, negative value is obtained). In this way the IMR has attempted to capture the joint effects of the unknown confounders in our study. Thus, to a certain extent, it mimics a variable which captures the aggregate effect of unmeasured predictors of treatment allocation. Therefore, if we adjust for this IMR in our main analysis looking at factors associated with the outcome variable then we should adjust for the unmeasured confounding and thus obtain unbiased estimates of the treatment effect.

7.2.4 Second stage model

We now fit a second stage model to assess the association between treatment allocation and treatment outcome. We again fit a regression model, but this time the dependent variable is the outcome (e.g. virological response, presence of an adverse event). Again, the theory suggests that the regression model should be fitted with a probit link function. The potential explanatory factors in this model are treatment allocation (regimen A or regimen B), the IMR and all factors which could potentially affect the outcome (O_1 to O_r and OT_1 to OT_t in the notation of subsection 4.2.1).

After obtaining the results of our second stage model, we first look at the parameter estimate and associated p-value for the IMR. If the parameter estimate suggests there is no association between the IMR and the outcome then Heckman's theory indicates that there is no unmeasured confounding present in our sample, and therefore we can use our standard methods to assess the association between treatment and response.

If we find that the IMR is associated with the outcome variable, then there is evidence that unmeasured confounding is present, and we must present our main study results from this second stage model. The parameter estimate of the treatment variable has now been adjusted for the unmeasured confounders, and thus, under the conditions set out in the Heckman modelling theory, should be an unbiased estimate of the treatment effect.

7.2.5 Potential limitations of Sample Selection models

When carrying out Sample Selection models, there are a number of potential limitations that must be investigated. Firstly, it is necessary to have information on at least one variable that is associated with treatment allocation only, or is associated with outcome only. We may encounter problems of co-linearity between variables included in the second stage model and the IMR – it is important to assess the correlation between each variable and the IMR to ensure that this is not an issue in each analysis performed. Another limitation of Sample Selection models occurs when there is insufficient information to estimate the probability of treatment allocation. Thus, we need information on at least one variable that is associated with this.

The theory behind Sample Selection models suggests that a probit link function should be used in both the first stage and the second stage model. Parameter estimates obtained from a probit model are not used frequently in the medical literature, and thus may not be as easy to interpret as results from a logistic model (odds ratios). Thus, we may not be left with a clear understanding of the magnitude of the treatment effect on the outcome. However, if the results from a logistic regression model are consistent with those produced from the probit regression model then we may present these results instead. I shall investigate this issue further in the data simulations in Section 7.4.

7.3 Simulation study to investigate usefulness of Sample Selection models

7.3.1 Introduction

Section 7.2 has described how Sample Selection models are derived. However, it is important to understand under which conditions, if any, Sample Selection models result in unbiased treatment estimates. This can only be investigated in a situation in which the true effect of treatment on response is known, and the amount of unmeasured confounding present is known. These issues can be addressed by carrying out a simulation study.

7.3.2 Methods

I first created a dataset with 10000 hypothetical individuals. For each individual, six variables were randomly generated: three normally distributed continuous variables

with zero mean and a standard deviation of one (T_c, O_c and OT_c) and three binary variables (T_b, O_b and OT_b) that take the values 1 or 0 with equal probability.

These six variables were split into three groups, and used to predict treatment allocation and/or the outcome variable in the following ways:

1) T_c and T_b predict treatment allocation only
2) O_c and O_b predict outcome only
3) OT_c and OT_b predict both treatment allocation and outcome (confounding variables)

Each variable was assumed to be associated with treatment or outcome as appropriate with a log odds ratio of 1.0 (i.e. an odds ratio of 2.7), except for the confounders (OT_c and OT_b) which were initially associated with treatment and outcome with a log odds ratio of 2.0 (i.e. an odds ratio of 7.4). It was assumed that treatment itself was not associated with outcome. Finally, a random error term with zero mean and a standard deviation of 1.7 was added to introduce a random element to treatment allocation, and a random error term with the same distribution was introduced to the outcome variable.

I began by investigating the situation when there was no unknown or unmeasured confounding present (Situation 0). In this situation all of T_c, T_b, O_c, O_b, OT_c and OT_b are assumed to be known and measured. Under these conditions, I investigated the results of five different models assessing the association between treatment and outcome. The five models used were:

(i) A standard logistic regression model adjusted for all known covariates
(ii) A propensity score model
(iii) A sample selection model with a probit link for both stages
(iv) A sample selection model with a logistic link for both stages
(v) A sample selection model with a probit link for the first stage model, and a logit link for the second stage model

My propensity score model consisted of a two-stage logistic regression model. In the first stage I included all known variables in a logistic regression model predicting treatment. From this logistic regression model I obtained the predicted score of being

allocated to treatment (the 'propensity score'). I then carried out a second logistic regression model investigating the association between treatment allocation and outcome, adjusting only for the propensity score.

For each of the five methodological approaches I noted the following information for each simulation run:

1) The parameter estimate (log odds ratio in logistic models, probit estimate in probit models) for treatment which was generated in the second stage model.
2) The p-value associated with the parameter estimate for treatment.
3) The parameter estimate associated with the IMR obtained from the second stage model (methods (iii)-(v)).
4) The p-value associated with the parameter estimate for the IMR obtained from the second stage model (methods (iii)-(v)).

I repeated the data simulation 1000 times and calculated summary statistics based on these repeated datasets.

After investigating the situation in which no unmeasured confounding was present, I went on to consider situations in which potential bias could occur. Firstly, I considered a situation identical to that described above for Situation 0, but this time I assumed that the continuous confounding variable, OT_c was unknown. Therefore, unmeasured confounding was now present. I then recorded the same information as described above and repeated the simulations 1000 times. I then assessed how well the five methodological approaches were able to provide unbiased estimates of the treatment effect and to identify the presence of unmeasured confounding under various scenarios. These scenarios are described in detail in Table 7.1. The rationale behind choosing them, and the interpretation of the results is explored in more detail in the results, Subsection 7.3.3.

Table 7.1 – Scenarios considered in simulation study to investigate effectiveness of Sample Selection models – all models assume no treatment effect on outcome

Variable Situation	Number individuals	Distribution of continuous variables	TC		TB		OC		OB		OTC			OTB			SD error term*
			Log OR for TA	Known	Log OR for TA	Known	Log OR for OV	Known	Log OR for OV	Known	Log OR for TA	Log OR for OV	Known	Log OR for TA	Log OR for OV	Known	
0	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	2.0	2.0	Y	2.0	2.0	Y	1.7
1	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	2.0	2.0	N	2.0	2.0	Y	1.7
2	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	2.0	2.0	Y	2.0	2.0	N	1.7
3	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	2.0	1.0	N	2.0	1.0	Y	1.7
4	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	2.0	1.0	Y	2.0	1.0	N	1.7
5	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	2.0	N	1.0	2.0	Y	1.7
6	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	2.0	Y	1.0	2.0	N	1.7
7	10000	Normal (0,1)	1.0	N	1.0	Y	1.0	Y	1.0	Y	2.0	2.0	N	2.0	2.0	Y	1.7
8	10000	Normal (0,1)	1.0	N	1.0	Y	1.0	Y	1.0	Y	2.0	2.0	Y	2.0	2.0	N	1.7
9	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	N	1.0	Y	2.0	2.0	N	2.0	2.0	Y	1.7
10	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	N	1.0	Y	2.0	2.0	Y	2.0	2.0	N	1.7
11	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.7
12	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.7
13	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.0
14	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.0
15	10000	Normal (0,1)	0.5	Y	0.5	Y	0.5	Y	0.5	Y	0.5	0.5	N	0.5	0.5	Y	1.7

Variable Situation	Number individuals	Distribution of continuous variables	TC		TB		OC		OB		OTC			OTB			SD error term*
			Log OR for TA	Known	Log OR for TA	Known	Log OR for OV	Known	Log OR for OV	Known	Log OR for TA	Log OR for OV	Known	Log OR for TA	Log OR for OV	Known	
16	10000	Normal (0,1)	0.5	Y	0.5	Y	0.5	Y	0.5	Y	0.5	0.5	Y	0.5	0.5	N	1.7
17	10000	Normal (0,1)	1.0#	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.7
18	10000	Normal (0,1)	1.0#	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.7
19	10000	Chi (1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.7
20	10000	Chi (1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.7
21	10000	Exp (2)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.7
22	10000	Exp (2)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.7
23	10000	Normal (0,1)	0.0	Y	0.0	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.7
24	10000	Normal (0,1)	0.0	Y	0.0	Y	1.0	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.7
25	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0 @	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.7
26	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0 @	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.7
27	10000	Normal (0,1)	0.05	Y	0.05	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.7
28	10000	Normal (0,1)	0.05	Y	0.05	Y	1.0	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.7
29	10000	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	N	1.7
30	500	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	N	1.0	1.0	Y	1.7
31	500	Normal (0,1)	1.0	Y	1.0	Y	1.0	Y	1.0	Y	1.0	1.0	Y	1.0	1.0	N	1.7

TA=treatment allocation; OV=outcome variable; OR=odds ratio; SD=standard deviation; *for random error term predicting treatment allocation, and for predicting outcome. # also associated with outcome with log odds ratio of 0.1- this is not known, and so T_c is thought only to be associated with treatment allocation. @ also associated with treatment allocation with log odds ratio of 0.1 – this is not known, and so O_c is thought only to be associated with outcome.

I then went on to investigate situations in which there was a true association between treatment and outcome. I again considered Situation 11 and Situation 12 as described in Table 7.1, but I assumed that there was an association between treatment and outcome, firstly with a log odds of +1.0 and secondly with a log odds of –1.0. I then performed 1000 simulations and employed the five methodological approaches described previously. The following information was noted for each simulation:

1) The parameter estimate (log odds in logistic models, probit estimate in probit models) for treatment obtained.
2) The p-value associated with the treatment estimate.
3) Whether the 95% CI of the estimate of the treatment effect contained the true treatment effect (methods (i), (ii), (iv) and (v))
4) The parameter estimate associated with the IMR (methods (ii)-(iv)).
5) The p-value associated with the parameter estimate for the IMR (methods (ii)-(iv))

7.3.3 Results

I first carried out my data simulations under the assumptions described in Situation 0 in Table 7.1. In this situation there are no unmeasured confounders, and thus one would expect all methodological approaches to give unbiased estimates of the treatment effect. The results from this situation are shown in Table 7.2.

When fitting a standard logistic regression model in this situation, it can be seen that the mean estimate of the treatment effect (here the log-odds ratio) obtained from the 1000 simulations was -0.01, with a range from -0.22 to +0.26. In 5.1% of simulations the associated p-value was less than 0.05. As there is no true treatment effect we would expect to obtain a mean of zero, and a p-value of less than 0.05 in 5% of simulations. Thus, when no unmeasured confounding is present a standard logistic regression model performs well. Similarly, the propensity scores model performed just as well in this situation. The mean (range) treatment effect observed was -0.03 (-0.18, +0.16) and the associated p-value was less than 0.05 in 3.9% of simulations.

Let us now consider the Sample Selection Models. As there is no unmeasured confounding present in this situation, one would expect the estimated association of the IMR with outcome obtained from the second stage model to have zero mean, and a p-value that was less than 0.05 in 5% of simulations. It can be seen from Table 7.2 that

this is the case for all three sample selection models. For, example, when a probit link function is used, the mean (range) parameter estimate is 0.00 (-0.30, +0.33) and a p-value of <0.05 was found in 5.2% of simulations. When considering the treatment effect, again we can see that all three Sample Selection Methods have given unbiased estimates of the treatment effect, as the mean estimate is very close to zero for each method. Furthermore, the p-value is less than 5% in approximately 5% of simulations. Therefore, in this simple situation in which there is no unmeasured confounding present, all methods perform well.

Table 7.2 – Results of simulation study when no unmeasured confounding is present and there is no true treatment effect

	Inverse Mills Ratio (IMR)		**Treatment effect**	
	Mean (range) estimate	% times p<0.05	Mean (range) estimate	% times p<0.05
Standard logistic regression	-	-	-0.01 (-0.22, +0.26)	5.1
Propensity Scores	-	-	-0.03 (-0.18, +0.16)	3.9
Sample Selection: probit link	0.00 (-0.30, +0.33)	5.2	0.00 (-0.23, +0.20)	5.1
Sample Selection: logit link	+0.03 (-0.19, +0.25)	6.9	-0.04 (-0.42, +0.38)	6.2
Sample Selection: probit link first stage, logit link second stage	-0.05 (-0.58, +0.53)	5.8	+0.03 (-0.37, +0.39)	6.2

Now let us consider a situation in which unmeasured confounding is present, starting with Situation 11 described in Table 7.1. Here, the continuous confounder, OT_c is assumed to be unknown or unmeasured, but again there is no true treatment effect. The results of this situation are shown in Table 7.3. Considering first the standard logistic regression model, the mean (range) estimate associated with the treatment effect is +0.73 (+0.53, +0.92) and in all simulations the p-value for the estimate was less than 0.05. Therefore, the logistic regression model consistently provided biased estimates of the treatment effect, and incorrectly described an association between treatment and outcome when in fact none existed. Similar results were seen when considering the propensity scores model.

When considering the Sample Selection Models with a probit link function for both stage models, the mean (range) estimate for the estimate associated with the IMR is +0.27 (+0.09, +0.42). This was statistically significant at the 5% level in 99.7% of simulations. Therefore, in nearly all simulations, the Sample Selection model correctly identified the presence of unmeasured confounding. When considering the treatment effect, the mean (range) estimate across the 1000 simulations was -0.01 (-0.25, +0.28) and the p-value was <0.05 in 5.0% of simulations (i.e. an unbiased treatment effect). Similar results were obtained when applying a Sample Selection model with a probit link function for the first stage model and a logit link function for the second stage model. However, the results were somewhat different when considering the Sample Selection model in which a logit link function was used for both stages. Here, the mean parameter estimate associated with the IMR and corresponding p-value correctly identified the presence of unmeasured confounding. However, the treatment estimate was still biased, with a mean (range) value of +0.21 (-0.09, +0.55) and a p-value of <0.05 in 66.9% of the simulations.

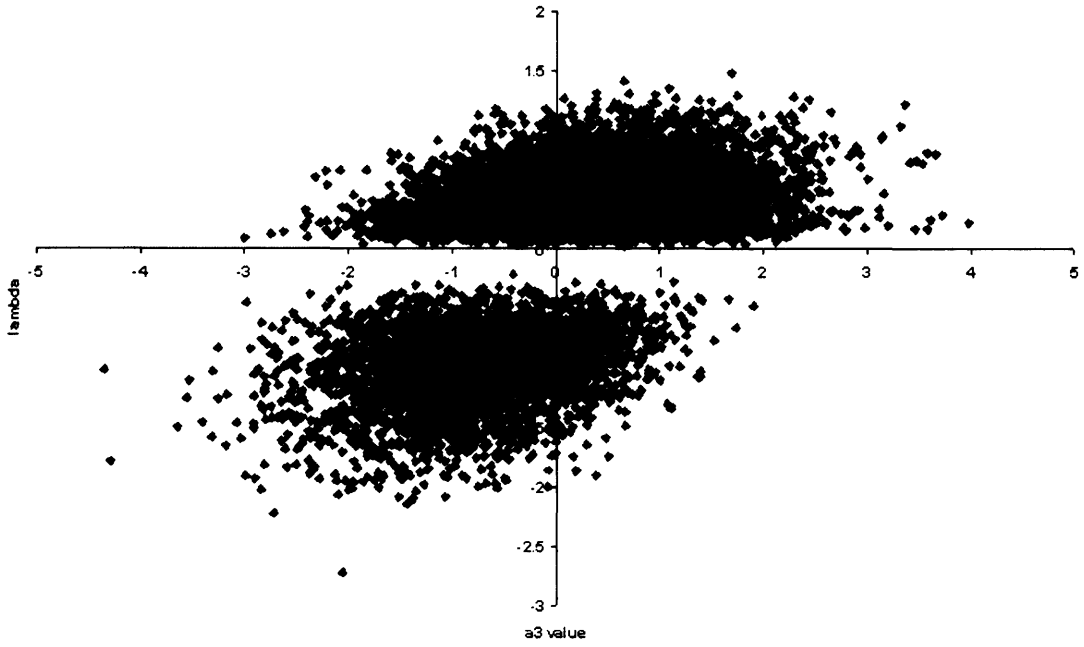
Table 7.3 – Results of simulation study when the continuous confounding variable (*OTc*) is unknown and there is no true treatment effect: Situation 11 described in Table 7.1

	Inverse Mills Ratio (IMR)		**Treatment effect**	
	Mean (range) estimate	% times p<0.05	Mean (range) estimate	% times p<0.05
Standard logistic regression	-	-	+0.73 (+0.53, +0.92)	100.0
Propensity Scores	-	-	+0.60 (+0.44, +0.75)	100.0
Sample Selection: probit link	+0.27 (+0.09, +0.42)	99.7	-0.01 (-0.25, +0.28)	5.0
Sample Selection: logit link	+0.28 (+0.09, +0.45)	99.7	+0.21 (-0.09, +0.55)	66.9
Sample Selection: probit link first stage, logit link second stage	+0.45 (+0.15, +0.17)	99.8	-0.02 (-0.42, +0.46)	5.3

I then wished to investigate whether the IMR obtained from the first stage Sample Selection model was in some way a "proxy" for the unmeasured confounder, OT_c. I therefore chose one simulation run from the situation described in Table 7.3, and plotted the IMR for each individual obtained from this particular run against that

individual's value for the unknown confounder OT_c to investigate their correlation. The results are shown in Figure 7.1. There is correlation between the two variables with a correlation coefficient of +0.54 ($p<0.0001$), which is perhaps not as close to one as we might expect if the sample selection model were "reconstructing" the unmeasured confounder through the IMR.

Figure 7.1 – Correlation between the IMR for each individual and the 'unobserved' confounding variable *OTc* in a single iteration of situation 11



Heckman model with a probit link function for both stages
(correlation coefficient=0.54; $p<0.0001$)

I next went on to consider the situation in which the binary confounder was assumed to be unknown (situation 12). The results of the 1000 simulations for this situation are shown in Table 7.4. It can be seen that, again, both the standard logistic regression and the propensity scores model gave biased estimates of the treatment effect. When considering the Sample Selection model with a probit link function for both the first stage model and the second stage model, the treatment estimate obtained (mean: 0.00; range -0.22, +0.29) was unbiased. However, the mean (range) parameter estimate associated with the IMR obtained from the 1000 simulations was small at +0.08 (-0.11, +0.22), and only in 37.6% of simulations was the corresponding p-values less than 0.05. Thus, the model was not always identifying the presence of

unmeasured confounding, although it was nonetheless able to adjust for this. Similar results were obtained from the other two Sample Selection models, although again the model with a logit link function for both stage models overestimated the proportion of times that the simulations found a p-value of less than 0.05 for the treatment effect approximately two-fold.

Table 7.4 – Results of simulation study when binary confounding variable (OT_b) unknown and there is no true treatment effect

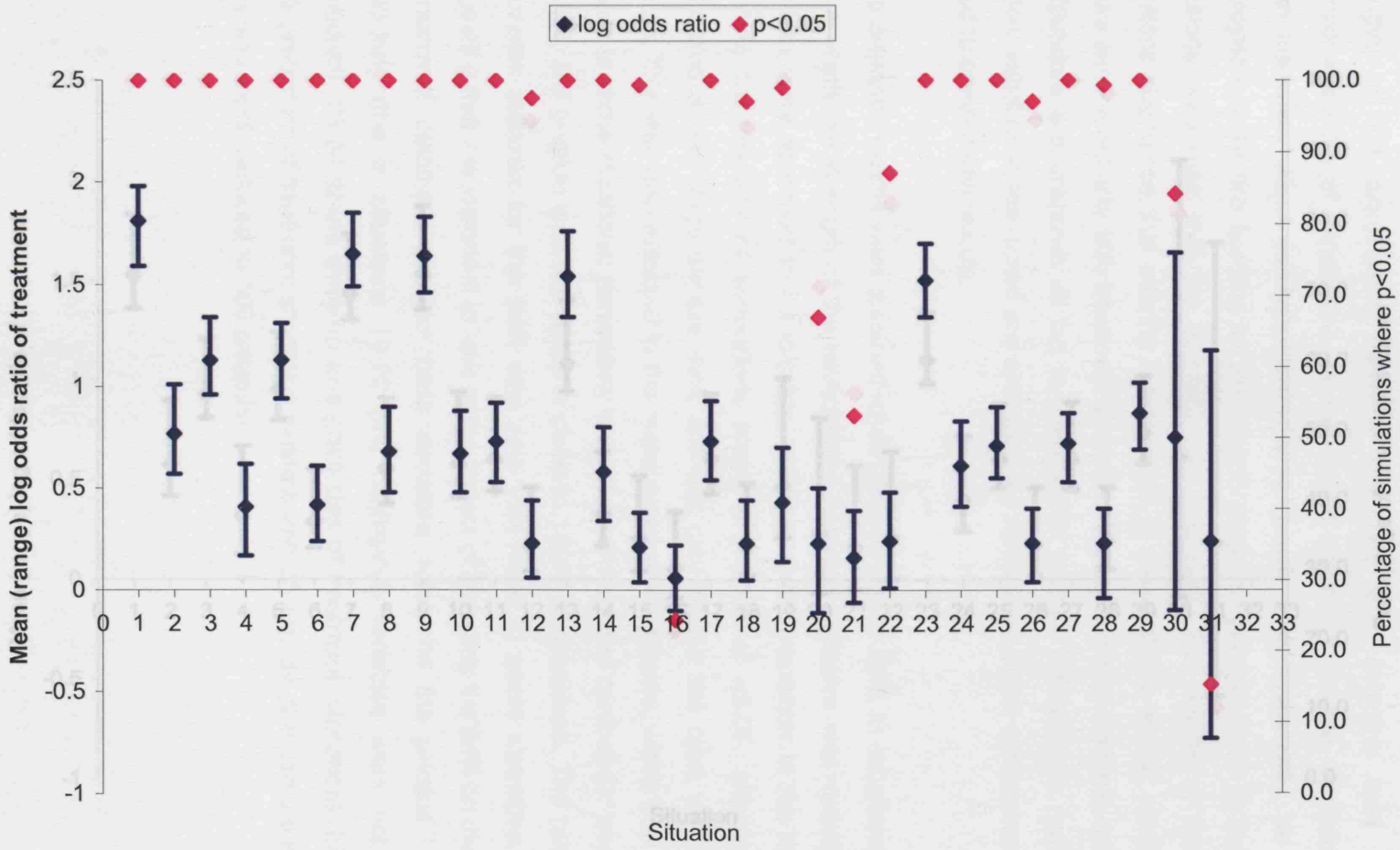
	Inverse Mills Ratio (IMR)		**Treatment effect**	
	Mean (range) estimate	% times p<0.05	Mean (range) estimate	% times p<0.05
Standard logistic regression	-	-	+0.23 (+0.06, +0.44)	97.5
Propensity Scores	-	-	+0.17 (+0.04, +0.33)	94.1
Sample Selection: probit link	+0.08 (-0.11, +0.22)	37.6	0.00 (-0.22, +0.29)	5.0
Sample Selection: logit link	+0.09 (-0.11, +0.23)	38.0	+0.07 (-0.23, +0.38)	10.9
Sample Selection: probit link first stage, logit link second stage	+0.14 (-0.18, +0.38)	39.6	0.00 (-0.37, +0.47)	5.2

I then went on to consider the other scenarios detailed in Table 7.1. The results of these simulations are shown in Figures 7.2-7.6, with each figure corresponding to one of the five methodological approaches. Figure 7.2 considers the results of the data simulations when using a standard logistic regression model. By first considering the percentage of times that the p-value was less than 0.05, it is immediately obvious that in all situations the model was finding a p<0.05 in most simulation runs when in fact there was no true different between treatment arms. Furthermore, in all simulations the mean estimate of the treatment effect obtained from the simulations (here equivalent to the log odds ratio) was not close to zero, the true treatment effect. The situations in which the least bias occurred was situation 16, in which the associations between variables was weakest (the association between all variables was a log odds ratio of 0.05, which is equivalent to an odds ratio of 1.64). Here, the mean (range) treatment parameter estimate was +0.06 (-0.10, +0.22), relatively close to zero, but a p-value of less than 0.05 was found in 24.4% of simulations, which is approximately five times more frequently than one would expect to see by chance. This implies that, if the

unmeasured confounding present is of the magnitude present in this example (which is still of a large magnitude) then there is less impact on the results of the standard logistic regression. However, in the comparable situation, situation 15, which was identical to situation 16 except that the unmeasured confounder was continuous rather than binary, a mean log odds ratio of +0.21 (+0.04, +0.38), and the associated p-value was <0.05 in 99.3% of simulation runs. Finally, in situation 34, in which the sample size was much reduced to 500 patients, a p-value of <0.05 was found in only 15.3% of simulations. However, the mean (range) log odds ratio for the treatment effect was +0.24 (-0.73, +1.18), indicating that although this effect was biased, there was a wide degree of sampling error in this situation.

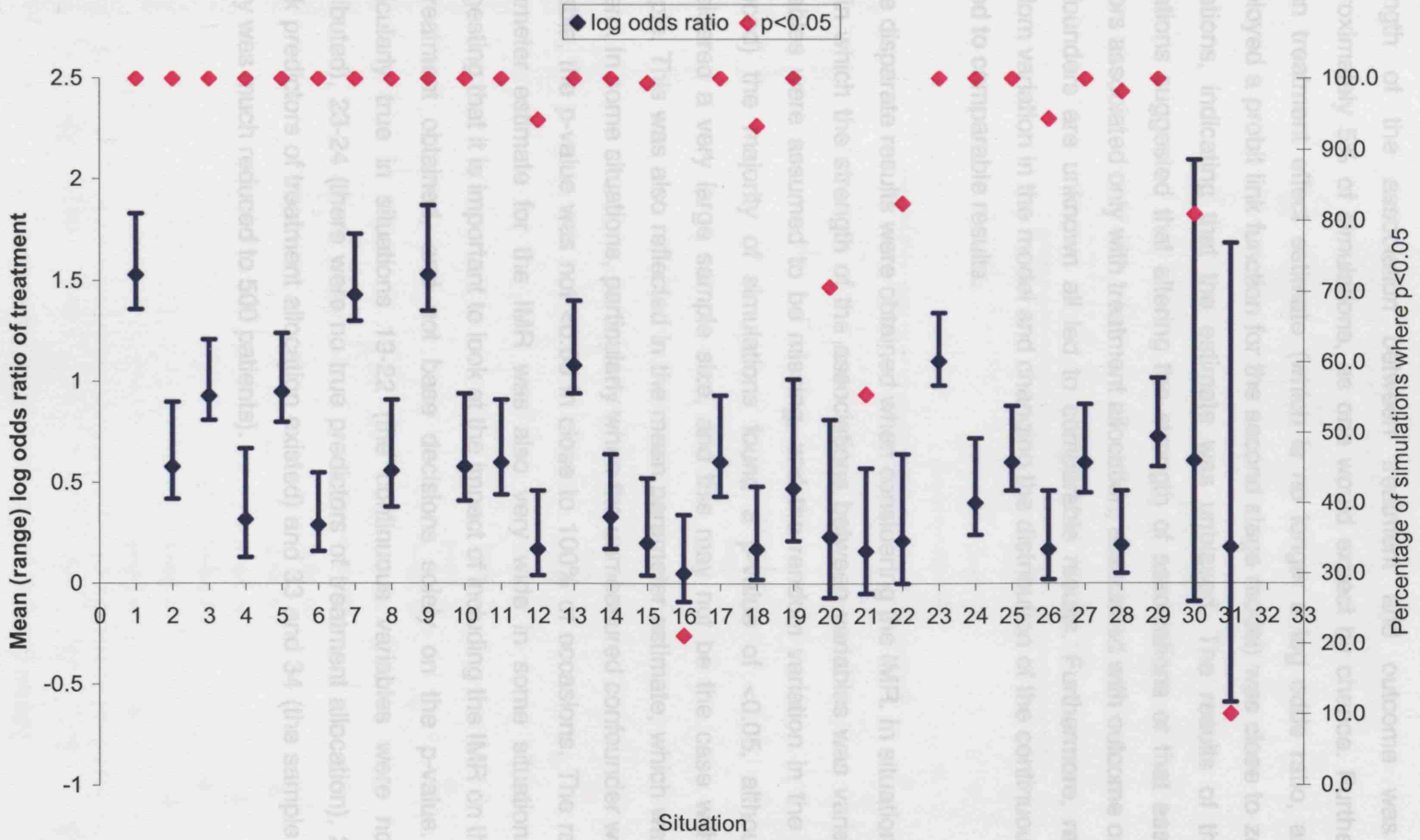
Figure 7.3 considers using a propensity scores model to investigate the association between treatment and outcome. The results are broadly similar to those observed in Figure 7.2. Again, in virtually all situations the percentage of simulation runs in which the p-value measuring the association between treatment and outcome was <0.05 was high, and often close to 100%. The mean treatment parameter estimate (log odds ratio) was not close to zero in most simulations, indicating that the treatment effect estimate was biased. The situation least affected by the bias caused by unmeasured confounding was again situation 16 (in which there was a weaker association between variables, and the OT_b was unknown). However, situation 15 which is identical to situation 16, except that the unknown confounder is continuous, led to biased results with a mean (range) log-odds ratio of +0.20 (+0.04, +0.36). In situation 34 (in which there was a much smaller sample size of 500 patients and OT_b was unknown), the p-value associated with the treatment effect being significant at the 5% level in 10% of simulation runs, likely reflecting the reduction in power. However, there was a wide range of log odds ratios observed, with a range of -0.09 to +1.40 (mean=+0.61).

Figure 7.2 – Results of simulation study using logistic regression model with no attempt to control for unmeasured confounding



For description of situations see Table 7.1

Figure 7.3 – Results of simulation study using propensity scores model with no attempt to control for unmeasured confounding



For description of situations see Table 7.1

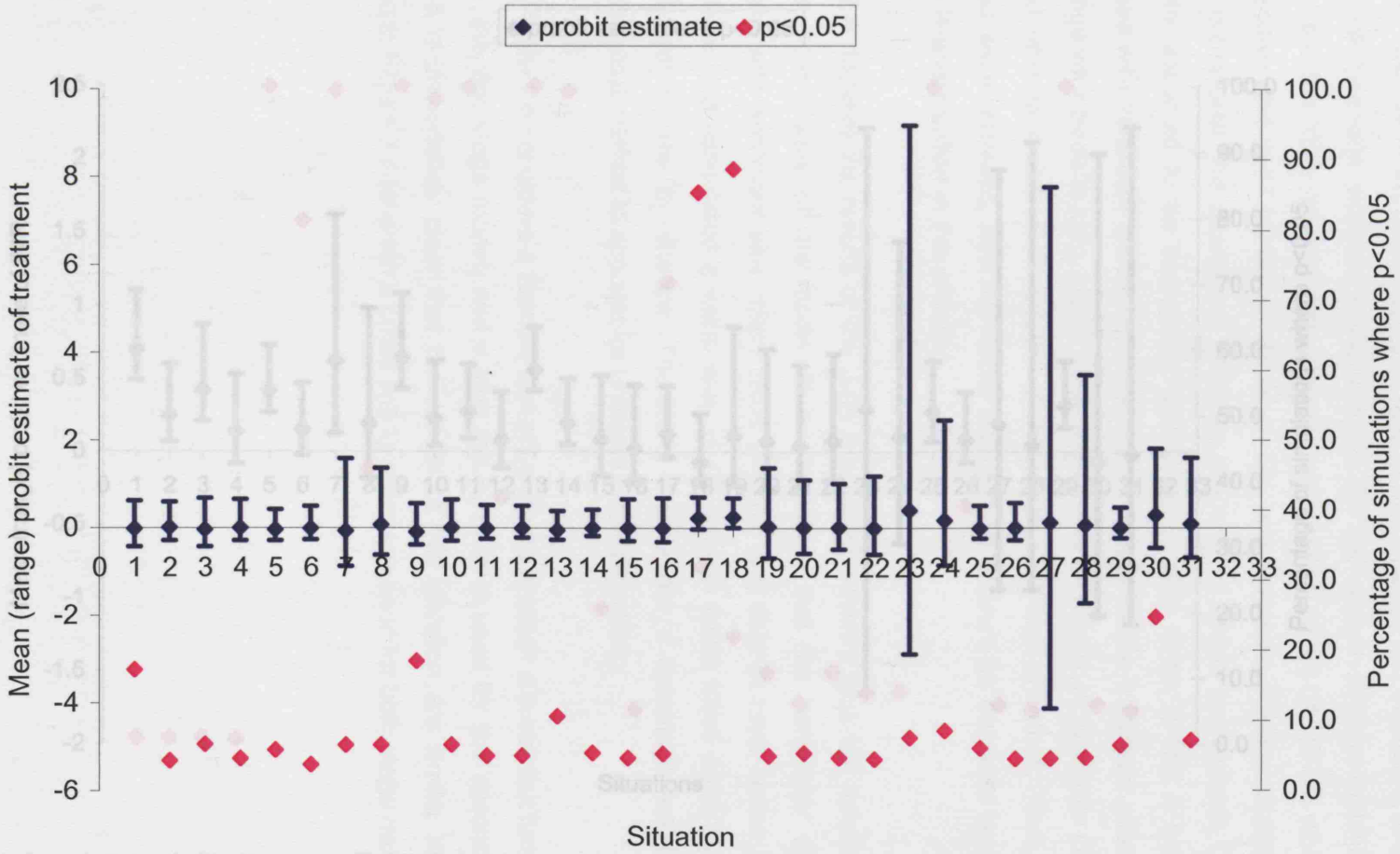
I next considered the first situation in which a Sample Selection model was fitted, using a probit link function for both stage models. Figure 7.4 contains information both on the obtained treatment effect estimate (Figure 7.4a) and the parameter estimate associated with the IMR obtained from the first stage model (Figure 7.4b).

First considering the treatment effect estimate, it is clear that, in most situations employed, the percentage of simulation runs in which the p-value assessing the strength of the association between treatment and outcome was <0.05 in approximately 5% of simulations, as one would expect by chance. Furthermore, the mean treatment effect estimate (which is no longer a log odds ratio, as we have employed a probit link function for the second stage model) was close to zero in most situations, indicating that the estimate was unbiased. The results of the different situations suggested that altering the strength of associations or that assuming that factors associated only with treatment allocation, associated with outcome only or other confounders are unknown all led to comparable results. Furthermore, reducing the random variation in the model and changing the distribution of the continuous variables all led to comparable results.

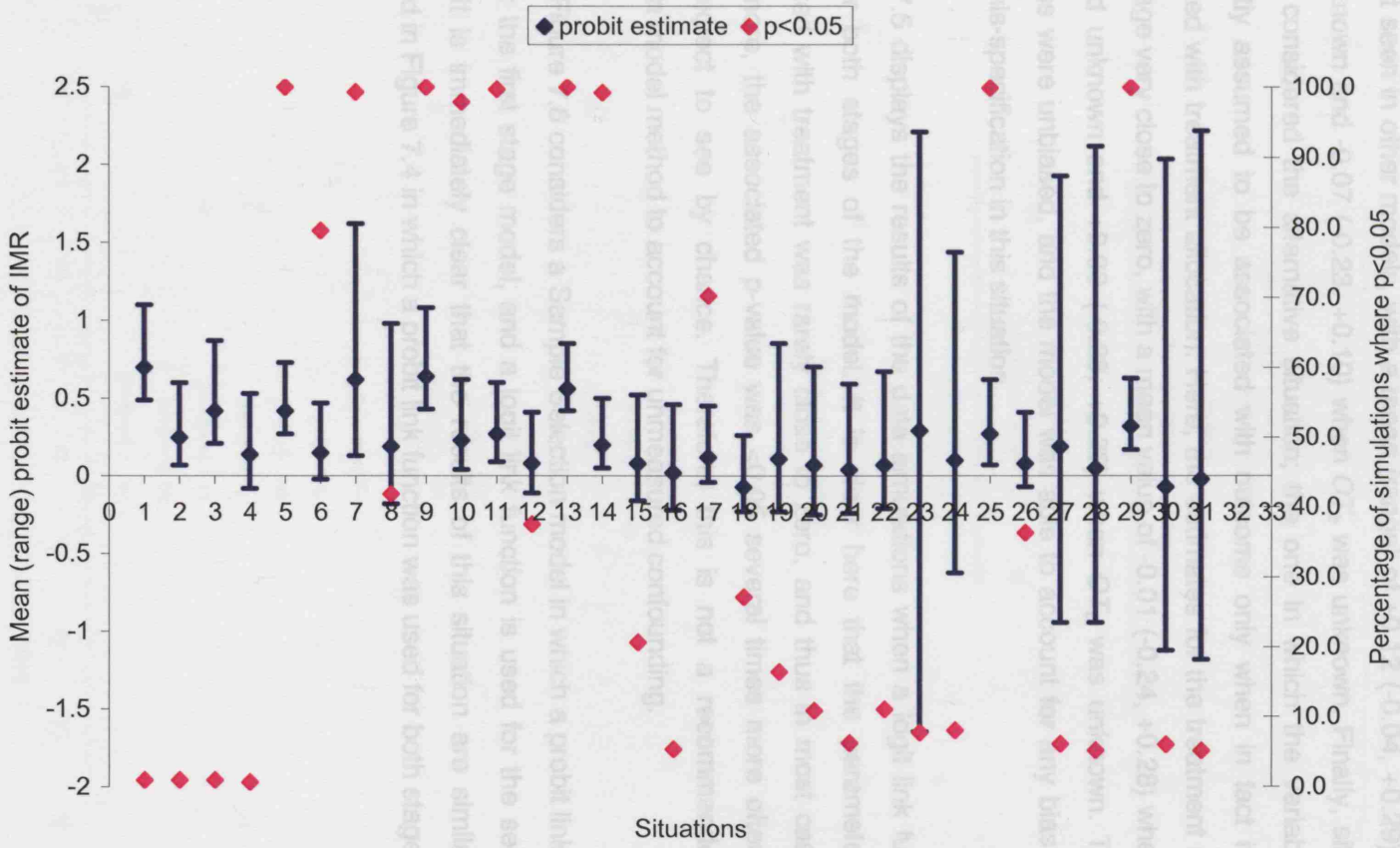
More disparate results were obtained when considering the IMR. In situations 1-16 and 31 (in which the strength of the associations between variables was varied, different variables were assumed to be missing, and the random variation in the model was reduced) the majority of simulations found a p-value of <0.05, although here I considered a very large sample size, and this may not be the case with a smaller sample. This was also reflected in the mean parameter estimate, which was not close to zero. In some situations, particularly when the unmeasured confounder was a binary variable, the p-value was not <0.05 in close to 100% of occasions. The range of the parameter estimate for the IMR was also very wide in some situations, perhaps suggesting that it is important to look at the impact of including the IMR on the estimate of treatment obtained, and not base decisions solely on the p-value. This was particularly true in situations 19-22 (the continuous variables were not normally distributed), 23-24 (there were no true predictors of treatment allocation), 27-28 (only weak predictors of treatment allocation existed) and 33 and 34 (the sample size of the study was much reduced to 500 patients).

Figure 7.4 – Results of simulation study using sample selection model with a probit link function

(a) Parameter estimates of association between treatment effect and outcome



(b) Parameter estimates of association between IMR and outcome



For description of situations see Table 7.1

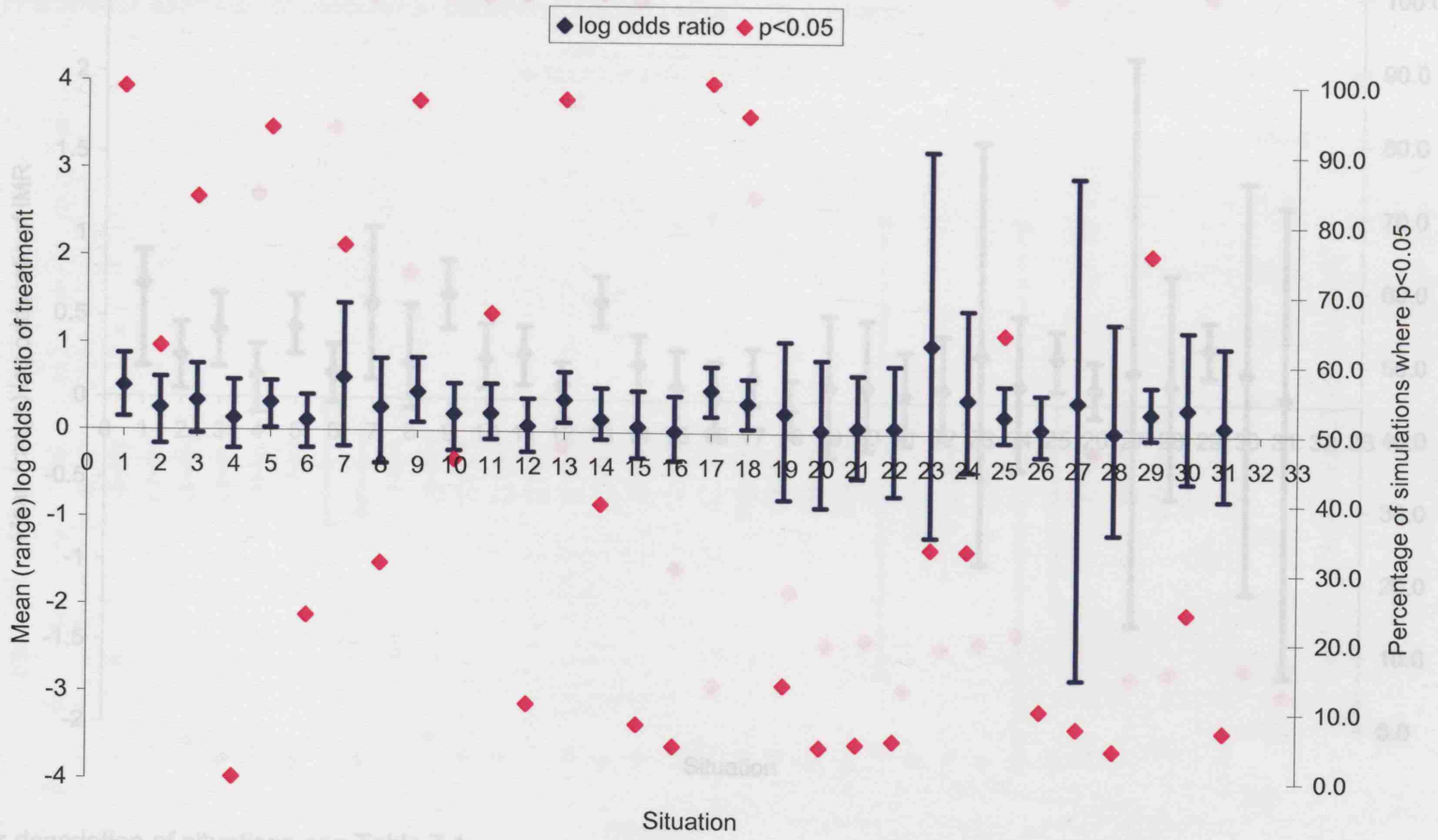
Four situations merit a special mention. In situations 17 and 18, it was assumed that T_c was incorrectly assumed to be associated with treatment allocation only when in fact it was weakly associated with outcome also. Thus, the second stage model was mis-specified. Here, the mean (range) parameter estimate associated with treatment was +0.23 (-0.02, +0.46) and the p-value was <0.05 in 85.3% of simulations when the OT_c was unknown (situation 17); the mean (range) parameter estimate associated with treatment was +0.25 (-0.01, +0.45) and the p-value was <0.05 in 88.6% of simulations when OT_b was unknown (situation 18). Thus, in this situation the treatment effect was biased. The parameter estimate associated with the IMR was on average much smaller than that seen in other models, with a mean (range) of +0.12 (-0.04, +0.29) when OT_c was unknown and -0.07 (-0.23, +0.10) when OT_b was unknown. Finally, situations 25 and 26 considered the alternative situation; the one in which the variable O_c was incorrectly assumed to be associated with outcome only when in fact it was also associated with treatment allocation. Here, the estimates for the treatment effect were on average very close to zero, with a mean value of -0.01 (-0.24, +0.28) when OT_c was assumed unknown and +0.00 (-0.28, +0.29) when OT_b was unknown. Thus, these estimates were unbiased, and the model was able to account for any bias caused by model mis-specification in this situation

Figure 7.5 displays the results of the data simulations when a logit link function was used for both stages of the model. It is clear here that the parameter estimate associated with treatment was rarely close to zero, and thus in most cases biased. Furthermore, the associated p-value was <0.05 several times more often than one would expect to see by chance. Therefore, this is not a recommended Sample Selection model method to account for unmeasured confounding.

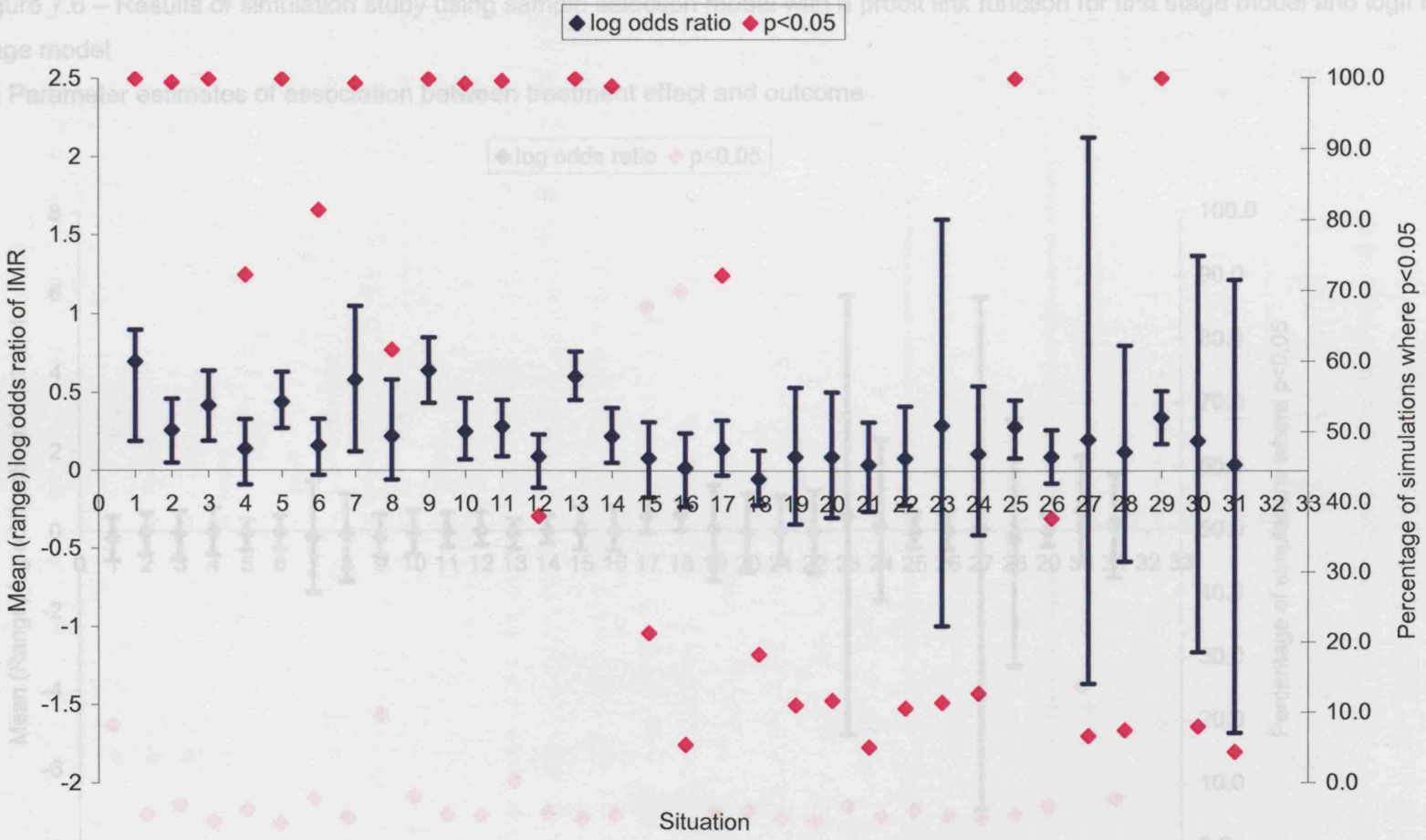
Finally, Figure 7.6 considers a Sample Selection model in which a probit link function is used for the first stage model, and a logit link function is used for the second stage model. It is immediately clear that the results of this situation are similar to those observed in Figure 7.4 in which a probit link function was used for both stage models.

Figure 7.5 – Results of simulation study using sample selection model with a logit link function

(a) Parameter estimates of association between treatment effect and outcome



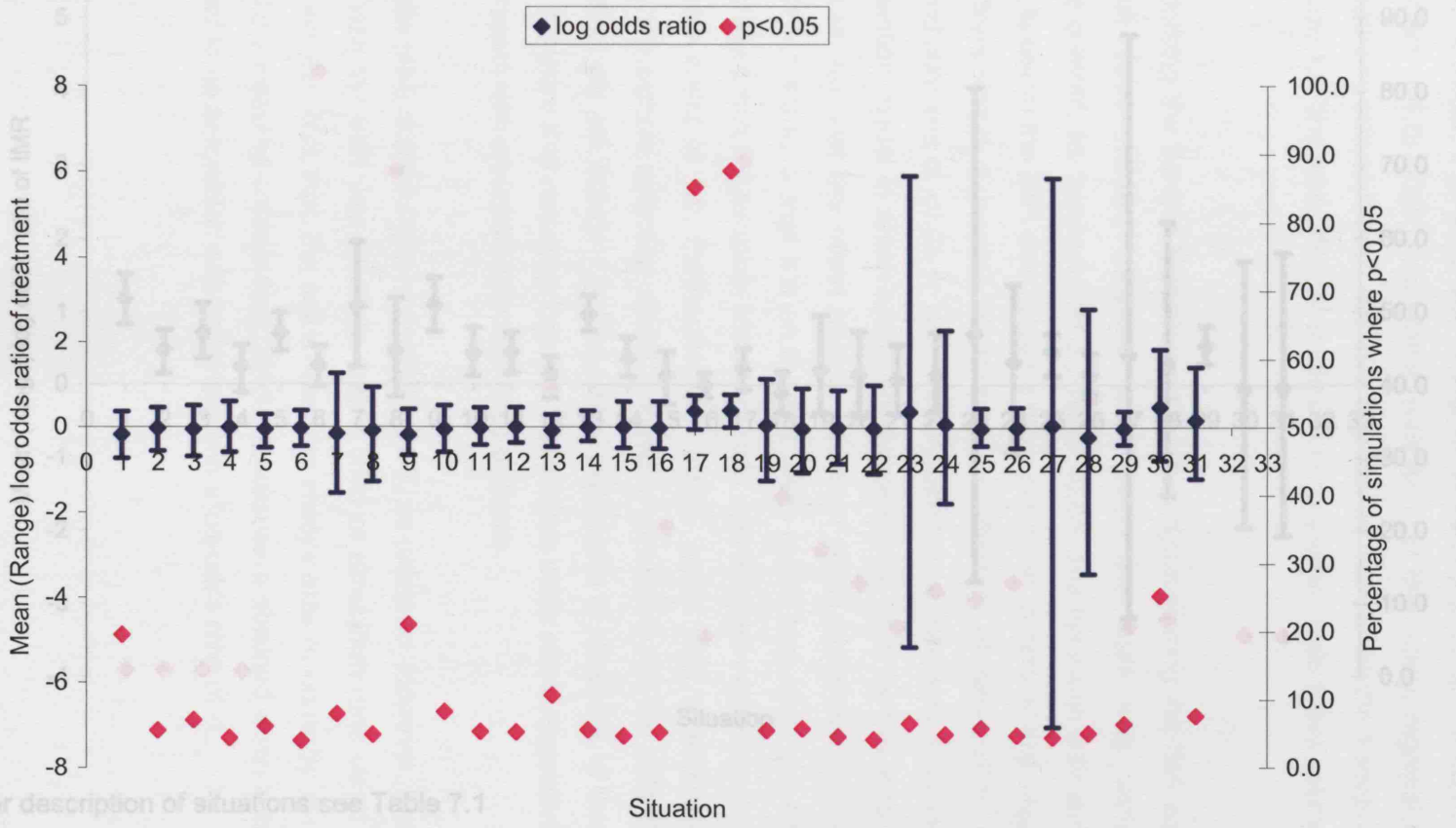
(b) Parameter estimates of association between IMR and outcome



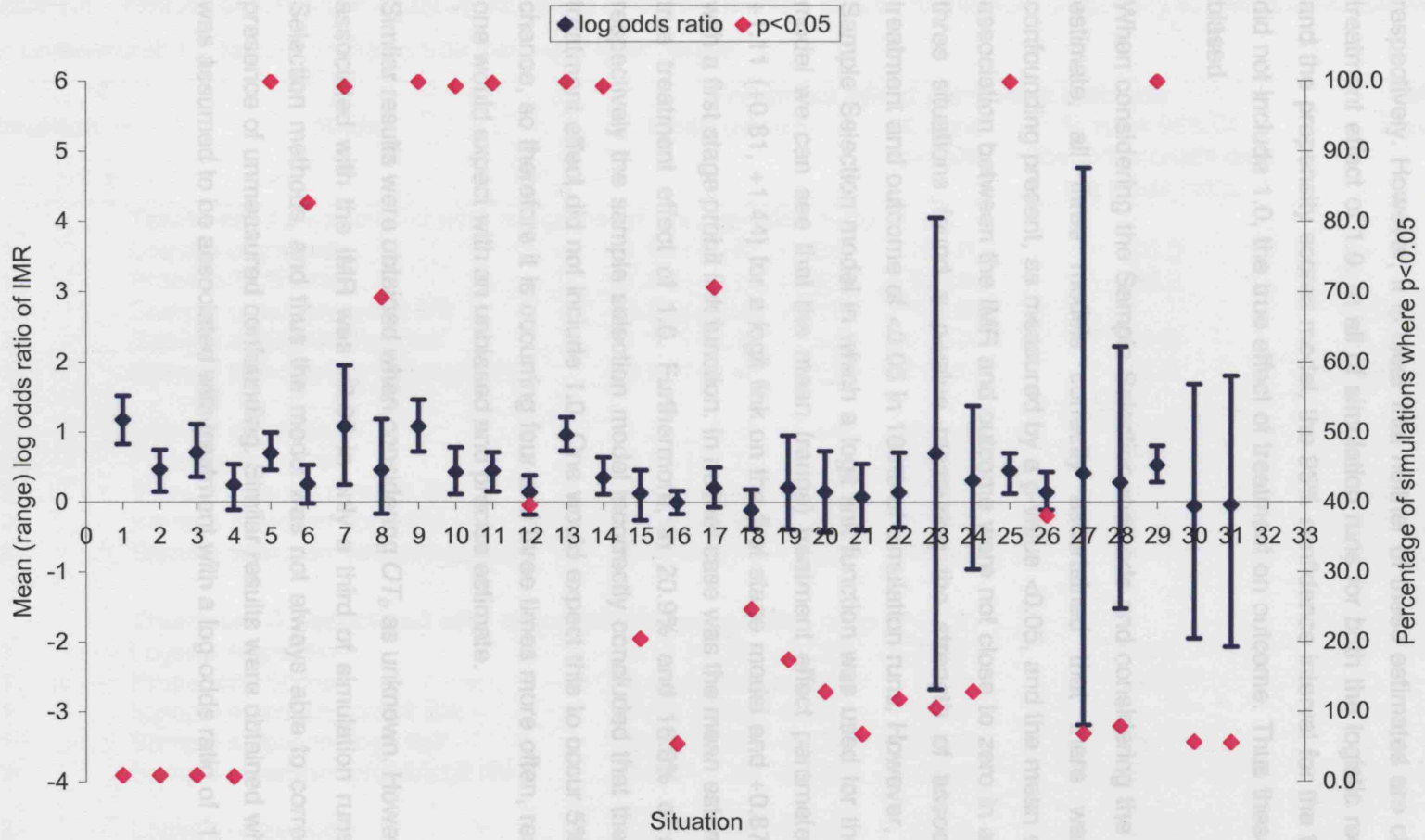
For description of situations see Table 7.1

Figure 7.6 – Results of simulation study using sample selection model with a probit link function for first stage model and logit link function for second stage model

(a) Parameter estimates of association between treatment effect and outcome



(b) Parameter estimates of association between IMR and outcome



For description of situations see Table 7.1

I next considered the situations in which there was a true treatment effect. The results of these simulations are shown in Table 7.5. I first assumed that treatment A was assumed with a higher prevalence of the outcome with a log-odds ratio of +1.0 (equivalent to an odds ratio of 2.7) and that OT_c was unknown. In the logistic regression and propensity scores models in this situation the models correctly found an association with treatment and outcome 100% of the time as the mean (range) treatment effect parameter estimate was +1.64 (+1.40, +1.83) and +1.36 (+1.18, +1.51) respectively. However, it is clear that neither of these estimates are close to the true treatment effect of 1.0. In all of simulation runs for both the logistic regression model and the propensity scores model, the 95% confidence interval for the treatment effect did not include 1.0, the true effect of treatment on outcome. Thus these estimates are biased

When considering the Sample Selection methods and considering the IMR parameter estimate, all three models correctly ascertained that there was unmeasured confounding present, as measured by a p-value <0.05, and the mean estimates of the association between the IMR and outcome were not close to zero in any situation. All three situations found a p-value measuring the strength of association between treatment and outcome of <0.05 in 100% of simulation runs. However, focusing on the Sample Selection model in which a logit link function was used for the second stage model we can see that the mean (range) treatment effect parameter estimate was +1.11 (+0.81, +1.44) for a logit link on the first stage model and +0.87 (+0.43, +1.31) with a first stage probit link function. In neither case was the mean estimate close to the true treatment effect of 1.0. Furthermore, in 20.9% and 16.8% of simulation runs respectively the sample selection model incorrectly concluded that the 95% CI of the treatment effect did not include 1.0. One would expect this to occur 5% of the time by chance, so therefore it is occurring four and three times more often, respectively, than one would expect with an unbiased and precise estimate.

Similar results were obtained when considering OT_b as unknown. However, the p-value associated with the IMR was <0.05 in only a third of simulation runs using Sample Selection methods, and thus the model was not always able to correctly identify the presence of unmeasured confounding. Similar results were obtained when treatment A was assumed to be associated with treatment with a log-odds ratio of -1.

Table 7.5 – Results of simulation study investigating ability of standard logistic regression, propensity scores and sample selection models to account for unmeasured confounding when a true treatment effect exists

Situation	Model	Treatment effect parameter estimate			IMR parameter estimate	
		Mean (range)	% runs p<0.05	% runs 95%CI doesn't include true log odds ratio	Mean (range)	% runs p<0.05
	Treatment A associated with outcome with a log-odds ratio of 1.0					
11	Logistic regression	+1.64 (+1.40, +1.83)	100.0	100.0	-	-
11	Propensity Scores	+1.36 (+1.18, +1.51)	100.0	100.0	-	-
11	Sample selection-probit link	+0.51 (+0.26, +0.77)	100.0	-	+0.27 (+0.11, +0.45)	100.0
11	Sample selection -logit link	+1.11 (+0.81, +1.44)	100.0	20.9	+0.28 (+0.10, +0.50)	99.8
11	Sample selection -probit/logit link	+0.87 (+0.43, +1.31)	100.0	16.8	+0.46 (+0.19, +0.78)	100.0
12	Logistic regression	+1.21 (+0.95, +1.44)	100.0	89.5	-	-
12	Propensity Scores	+0.93 (+0.74, +1.10)	100.0	20.5	-	-
12	Sample selection -probit link	+0.57 (+0.30, +0.89)	100.0	-	+0.08 (-0.11, +0.27)	32.6
12	Sample selection -logit link	+1.04 (+0.66, +1.44)	100.0	6.7	+0.09 (-0.21, +0.29)	36.5
12	Sample selection -probit/logit link	+0.97 (+0.51, +1.52)	100.0	5.2	+0.14 (-0.18. +0.48)	35.6
	Treatment A associated with outcome with a log-odds ratio of –1.0					
11	Logistic regression	-0.14 (-0.28, +0.02)	75.2	100.0	-	-
11	Propensity Scores	-0.11 (-0.24, +0.02)	67.8	100.0	-	-
11	Sample selection -probit link	-0.52 (-0.80, -0.29)	100.0	-	+0.26 (+0.10, +0.43)	100.0
11	Sample selection -logit link	-0.64 (-0.99, -0.37)	100.0	98.1	+0.28 (-0.11, +0.47)	100.0
11	Sample selection -probit/logit link	-0.85 (-1.32, -0.48)	100.0	23.1	+0.43 (+0.18, +0.71)	100.0
12	Logistic regression	-0.73 (-0.88, -0.54)	100.0	100.0	-	-
12	Propensity Scores	-0.55 (-0.68, -0.41)	100.0	100.0	-	-
12	Sample selection -probit link	-0.56 (-0.78, -0.33)	100.0	-	+0.08 (-0.07, +0.22)	35.9
12	Sample selection -logit link	-0.88 (-1.16, -0.60)	100.0	27.9	+0.08 (-0.08, +0.25)	34.4
12	Sample selection -probit/logit link	-0.94 (-1.31, -0.55)	100.0	8.8	+0.13 (-0.13, +0.38)	35.9

For definitions of situations see Table 7.1

7.3.4 Summary

The results of this data simulation study have shown that standard logistic regression models and propensity score models do not perform well when unmeasured confounding is present in studies. The estimates of the treatment effect in practically all situations were biased, and the associated p-value was frequently <0.05. Similarly, a Sample Selection model in which a logit link function was used for both stage models led to biased results. In contrast, the use of a Sample Selection model with a probit link function for the first stage model, and either a probit or logit link function for the second model generally, with a few exceptions, led to unbiased treatment estimates. Although a probit link function should be used under Heckman theory, using a second stage model with a logit link function has the advantage that the treatment effect estimates are log odds ratios, and thus can be exponentiated to obtain odds ratios. Thus, if the results of this model are similar to those observed when using a probit link function for both stage models, it is probably reasonable to use a logit link function for the second stage model.

7.4 Application of Sample Selection models to compare the occurrence of hypercholesterolaemia amongst those receiving efavirenz- and nevirapine-containing regimens

7.4.1 Introduction

I shall now apply Sample Selection models to a real-life example to investigate how they might be used in practice. As described in Section 7.1, it is of interest to compare different antiretrovirals to investigate whether they are associated with different toxicity profiles as this could potentially limit the ability of patients to adhere to life-long treatment. I have chosen to compare the most commonly prescribed PI/NNRTI in recent calendar years; EFV and LPV.

7.4.2 Methods

Patients included in this study were from the Royal Free cohort (described in Chapter 3). Individuals included in this study were previously antiretroviral-naïve individuals who started HAART consisting of at least three drugs including LPV or EFV (but not both) from 1st January 2001 onwards. Included individuals were also required to have a pre-

treatment CD4 cell count and HIV RNA viral load measurement in the six month period before starting ART, and at least one viral load or CD4 cell measurement after starting treatment. Individuals were followed from the date of starting cART. Their date of last follow-up was defined as the date of their last recorded CD4 cell count or viral load measurement. I considered the percentage of individuals discontinuing LPV or EFV in the first 12 weeks of treatment. Any individual with less than 8 weeks of follow-up was assumed to have discontinued all treatment by 12 weeks to account for those lost to follow-up. Individuals making dose alterations to their regimens, and switching to other antiretrovirals were considered also to have discontinued the drug, as these switches may also be toxicity related.

A comparison of the tolerability of LPV and EFV was then conducted. I began by comparing the two regimens using a standard logistic regression model. Potential confounding factors that were adjusted for in the model were calendar year of starting HAART, pre-HAART CD4 cell count, sex and risk group (combined into a composite variable as there is high co-linearity between these variables), age at HAART, pre-HAART viral load and NRTI backbone received.

I then compared the tolerability of LPV and EFV using sample selection models to account for unknown and unmeasured confounders, such as depression and clinician choice. I chose the following predictors of treatment allocation (T), of outcome (O) and of both treatment choice anc outcome (OT):

- Calendar year (OT)
- Pre-cART CD4 cell count (OT)
- Sex/risk group (OT)
- Age (OT)
- Viral load (T)
- NRTI backbone (O)

Calendar year was thought to be associated with both treatment choice and outcome as LPV and EFV were licensed at different time points (LPV in 2000 and EFV in 1998), and thus EFV is likely to have been prescribed at earlier time points, although I only included those starting HAART from 2001 onwards to attempt to minimise these biases. Calendar year is also likely to be associated with treatment discontinuation, as clinicians gain more experience of managing toxicities. Pre-HAART CD4 cell count was also included as a potential confounder. There is some evidence that there is an association between the CD4 count at the time of starting HAART, and the choice of

first-line regimen [361]. There is also some evidence [119;375-379] that the pre-HAART CD4 cell count is associated with drug-related toxicity, although this has not been shown in all studies including Chapter 6 in this thesis [199;213;333;380;381]. Therefore, given the dangers of incorrectly assuming a confounder is only associated with treatment choice (see situations 17 and 18 in data simulation study) I felt that it was better to consider this as a potential confounder. Sex and risk group were considered as potential confounders, as EFV has been shown to be associated with teratogenic effects in pregnancy [403], and there is also evidence of differential treatment discontinuance rates in men and women [404]. Finally, age was also considered to be a potential confounder as age has been shown to be associated with treatment choice and drug tolerability [166].

Pre cART viral load was considered as only being associated with treatment choice, as high pre-cART viral loads might influence treatment choice, but it was thought that these are unlikely to be associated with treatment discontinuation. The NRTI backbone was considered to be only associated with outcome, as it is likely that the clinician would first choose the PI or NNRTI for the regimen, and then subsequently choose the accompanying NRTIs. However, as NRTIs are associated with a number of toxicities, these may well influence treatment discontinuations.

Sample selection models were implemented using a probit link function for the first stage model, and both with a probit link function and with a logit link function for the second stage model. Analyses were carried out using SAS version 9.1. (SAS Institute Inc, Cary, NC, USA).

7.4.3 Results

There were a total of 595 individuals included in the study. Two hundred and sixty four (264; 44.4%) started HAART with LPV and 331 (55.6%) individuals started EFV. The characteristics of these individuals at the time of starting HAART are shown in Table 7.6. There were differences in the year of starting HAART, with a higher percentage of those starting EFV doing so in 2001 (64 [19.3%] started EFV compared to only 30 [11.4%] starting LPV; p=0.004). There were also differences in the ART regimen started. Although for both groups the majority started three antiretrovirals (when not counting ritonavir which is administered to boost lopinavir levels), 256 (97.0%) of those starting LPV did so as part of a three drug regimen compared to 296 (89.4%) of those starting EFV. Furthermore, 195 (58.9%) of those starting EFV used a backbone consisting of AZT+3TC compared to 125 (47.4%) of those starting LPV. There were

few differences in pre-HAART viral loads, but those starting EFV on average had higher pre-HAART CD4 cell counts (median 209 cells/mm^3 for EFV compared to 181 cells/mm^3 for LPV; p=0.01) and were younger (median 35 years for EFV compared to 38 years for LPV; p=0.001).

Table 7.6 – Characteristics of participants starting first-line HAART with lopinavir or efavirenz in the Royal Free cohort

		EFV-containing HAART	LPV-containing HAART	p-value
Number		331 (100.0%)	264 (100.0%)	
Year of starting cART	2001	64 (19.3%)	30 (11.4%)	0.004
	2002	75 (22.7%)	66 (25.0%)	
	2003	78 (23.6%)	52 (19.7%)	
	2004	76 (23.0%)	59 (22.4%)	
	2005	37 (11.2%)	54 (20.5%)	
	2006	1 (0.3%)	3 (1.1%)	
Number of drugs in regimen*	3	296 (89.4%)	256 (97.0%)	0.0004
	4 or 5	35 (10.6%)	8 (3.0%)	
NRTI backbone	AZT+3TC	195 (58.9%)	125 (47.4%)	0.001
	TDF+3TC	37 (11.2%)	45 (17.1%)	
	D4T+3TC	18 (5.4%)	33 (12.5%)	
	Other	81 (24.5%)	61 (23.1%)	
Gender / risk group	Male/ homosexual	184 (55.6%)	141 (53.4%)	0.26
	Male/ other risk	76 (23.0%)	52 (19.7%)	
	Female	71 (21.5%)	71 (26.9%)	
Pre-HAART viral load (log_{10} copies/ml)	Median (IQR)	5.2 (4.6, 5.6)	5.1 (4.8, 5.6)	0.81
Pre-HAART CD4 cell count (cells/mm^3)	Median (IQR)	209 (78, 362)	181 (90, 257)	0.01
Age (years)	Median (IQR)	35 (31, 42)	38 (33, 44)	0.001

LPV=lopinavir; EFV=efavirenz; HAART=highly active antiretroviral therapy; NRTI=nucleoside reverse transcriptase inhibitors; AZT=zidovudine; 3TC=lamivudine; TDF=tenofovir; D4T=stavudine; IQR=inter quartile range
*not including ritonavir as used to boost lopinavir levels. P-values calculated using chi-squared tests and Mann Whitney U tests as appropriate.

I considered the percentage discontinuing EFV or LPV within 12 weeks of starting treatment. Eleven (3.3%) individuals starting EFV and 9 (3.4%) individuals starting LPV had less than 8 weeks follow-up and thus were considered to have discontinued EFV/LPV before 12 weeks. In total 59 (17.8%) participants starting EFV and 28

(10.6%) starting LPV had discontinued the drug by 12 weeks (p=0.01; chi squared test). In a standard logistic regression model (Table 7.7), this was associated with an odds ratio of 1.83 (95% confidence interval [CI] 1.13, 2.96; p=0.01), indicating that those receiving EFV had an 83% higher odds of discontinuation of the drug compared to those receiving LPV. After adjustment for other factors this odds ratio estimate became 2.42 (95% CI 1.41, 4.15; p=0.001). Thus, using standard techniques, there is evidence that EFV is associated with greater short-term toxicity as measured by discontinuing the drug. Other factors associated with treatment discontinuation were later calendar years (odds ratio [OR]=1.39 per year later; 95% CI 1.11, 1.75; p=0.004) and pre-HAART viral load (OR=0.98 per 1 log higher; 95% CI 0.70, 1.37; p=0.01).

Table 7.7 – Results of logistic regression model investigating factors associated with LPV/EFV discontinuation within 12 weeks of starting treatment

		Univariable analysis			Multivariable analysis		
		OR	95% CI	p	OR	95% CI	p
Antiretroviral	*EFV*	1.83	1.13, 2.96	0.01	2.42	1.41, 4.15	0.001
	LPV	1.00	-		1.00	-	
Year of cART	*Per year later*	1.23	1.03, 1.46	0.02	1.39	1.11, 1.75	0.004
Accompanying NRTIs	*AZT+3TC*	1.29	0.70, 2.37	0.13	1.64	0.82, 3.29	0.14
	TDF+3TC	2.38	1.14, 4.93		2.26	1.05, 4.89	
	D4T+3TC	0.48	0.48, 3.25		2.67	0.89, 8.00	
	Other	1.00	-		1.00	-	
Gender/risk group	*Male/homosexual*	0.44	0.26, 0.74	0.002	0.47	0.26 0.83	0.009
	Male/other	0.56	0.30, 1.06	0.08	0.56	0.29, 1.09	0.09
	Female	1.00	-		1.00	-	
Pre-HAART viral load	*Per 1 log higher*	0.91	0.67, 1.23	0.53	0.98	0.70, 1.37	0.91
Pre-HAART CD4 count	*Per 100 cells higher*	0.86	0.74, 1.01	0.06	0.91	0.76, 1.09	0.31
Age	*Per 10 years older*	1.03	0.80, 1.33	0.82	1.02	0.76, 1.34	0.89

LPV= lopinavir; EFV= efavirenz; HAART= highly active antiretroviral therapy; NRTI= nucleoside reverse transcriptase inhibitors; AZT= zidovudine; 3TC= lamivudine; TDF= tenofovir; D4T= stavudine; IQR= inter quartile range; OR= odds ratio; CI= confidence interval

I shall now compare the short-term tolerability of EFV and LPV using sample selection models to account for unmeasured confounders. I begin by carrying out a first stage model in which the outcome is whether the patient received LPV or EFV, and the explanatory variables are those listed in Subsection 7.5.2. The results of the first stage

model are shown in Table 7.8. It can be seen that earlier years of starting HAART, males with a homosexual risk for HIV infection, those with lower pre-HAART viral loads, those with lower pre-HAART CD4 cell counts, and those of older age were more likely to start a first-line HAART regimen containing EFV. The IMRs created from the first stage model took values in the range +0.19 to +1.79 for those receiving EFV with a median value of +0.65 and those receiving LPV had IMRs in the range -1.68 to -0.31 with a median value of -0.82. I also considered the correlation between the variables included in the second stage model and the IMR. There was no evidence of any correlation, except for a correlation with the NRTI backbone d4T/3TC (r=-0.15; p=0.0005).

Table 7.8 – Results of first stage sample selection model investigating predictors of allocation to EFV (rather than LPV) using a probit link function

		Estimate	95% CI	p
Year of HAART	*Per year later*	-0.16	-0.24, -0.08	<0.0001
Gender/risk group	*Male/homosexual*	+0.31	+0.04, +0.58	0.02
	Male/other	+0.12	-0.20, +0.43	0.46
	Female	0.00	-	
Pre-HAART viral load	*Per 1 log higher*	-0.17	-0.33, -0.02	0.03
Pre-HAART CD4 count	*Per 100 cells higher*	-0.14	-0.21, -0.08	<0.0001
Age	*Per 10 years older*	+0.12	+0.00, +0.24	0.04

LPV= lopinavir; EFV= efavirenz; cART= combination antiretroviral therapy; NRTI= nucleoside reverse transcriptase inhibitors; AZT= zidovudine; 3TC= lamivudine; TDF= tenofovir; D4T= stavudine; IQR= inter quartile range; OR= odds ratio; CI= confidence interval

I then carried out a second stage model investigating factors associated with discontinuing EFV or LPV within 12 weeks of starting cART. The results of this model when using a probit link function and a logit link function are shown in Table 7.9. In both cases the parameter estimate for the IMR was not associated with the outcome, and the parameter estimate was reasonably close to zero (estimate= –0.02) in the probit model, and the odds ratio in the logit model was close to one (OR=1.14). However, in both models the accompanying 95% CI was wide, perhaps reflecting the small number of individuals experiencing an event in the study (59 patients in the EFV group and 28 LPV group). When comparing the parameter estimates obtained from the

second stage sample selection model with a logit link for the treatment effect with that compared to the standard regression model from Table 7.8, in which I observed an OR of 2.42. The odds ratio obtained in the sample selection model was 1.97 (95% CI 0.01, 551; p=0.81) and, although this estimate is not precise, the actual estimates are relatively similar.

Table 7.9 – Results of second stage sample selection model investigating factors associated with LPV/EFV discontinuation within 12 weeks of starting treatment

		Probit link function			**Logit link function**		
		Estimate	95% CI	p	OR	95% CI	p
Antiretroviral	*EFV*	+0.38	-2.51, +3.27	0.80	1.99	0.01, 360.6	0.80
	LPV	+0.00	-		0.00	-	
IMR		+0.06	-1.73, +1.85	0.95	1.13	0.05, 28.33	0.94
Year	*/ year later*	+0.17	-0.03, +0.37	0.09	1.38	0.96, 1.98	0.08
NRTIs	*AZT+3TC*	+0.27	-0.10, +0.64	0.16	1.64	0.82, 3.30	0.16
	TDF+3TC	+0.45	+0.02, +0.87	0.04	2.26	1.04, 4.88	0.04
	D4T+3TC	+0.52	-0.06, +1.10	0.08	2.68	0.90, 8.03	0.08
	Other	+0.00	-		1.00	-	
Gender/risk group	*Male/homosexual*	-0.41	-0.84, +0.01	0.05	0.47	0.22, 1.01	0.05
	Male/other	-0.34	-0.73, +0.06	0.09	0.56	0.27, 1.22	0.15
	Female	+0.00	-		1.00	-	
Pre-HAART CD4 count	*/100 cells higher*	-0.05	-0.21, +0.11	0.53	0.91	0.68, 1.22	0.51
Age	*/10 years older*	+0.01	-0.19, +0.21	0.90	1.03	0.71, 1.49	0.88

LPV= lopinavir; EFV= efavirenz; HAART= highly active antiretroviral therapy; NRTI= nucleoside reverse transcriptase inhibitors; AZT= zidovudine; 3TC= lamivudine; TDF= tenofovir; D4T= stavudine; IQR= inter quartile range; OR= odds ratio; CI= confidence interval; IMR= Inverse Mills Ratio

7.5 Discussion

As discussed in Chapter 2, the most appropriate setting in which to carry out unbiased treatment comparisons is in an RCT. However, for practical and ethical reasons, this is not always possible. One of the major limitations of observational studies is that one can never rule out the presence of unmeasured confounding in this setting. Thus, sample selection models are an attractive option. However, in order to apply them in the context of unmeasured confounding, it is important to ensure that they will provide unbiased estimates of treatment effects.

I set out to consider a number of data simulations that would be comparable to the situations in which we might wish to use sample selection models. The results of my data simulations show that sample selection models give unbiased estimates in several simple situations. However, there are a number of scenarios in which the models led to biased results. Firstly, sample selection models are based on the assumption of a bivariate normal distribution, and hence the use of a probit link function is required. Although it was perhaps not expected, the results of my simulations indicate that the use of a logit link function, which follows a log normal distribution, in the first stage model leads to biased results in the situations considered. As the use of a probit link function for the first stage model is always correct, it seems sensible to always do so. However, a recent review of the use of sample selection models in the Criminology field found that a logit link function was used for the first stage model in three quarters of studies [405]. In contrast, I found that the choice of a probit or logit link function for the second stage model was less important in the situations considered. As we might wish to use a logit link function here to obtain odds ratios, it seems reasonable to compare the results of analyses using both link functions to ensure they lead to consistent results, and then choose to use a logit link function if this is the case.

I also found that it was important to correctly specify which variables were associated with treatment, which were associated with outcome, and which were associated with both. Although incorrectly specifying a confounding variable as being associated only with outcome did not lead to biased results, mis-specifying a confounder as being associated only with treatment allocation meant that the sample selection models no longer gave correct answers. This is perhaps to be expected, as the IMR is calculated based on the results of the first stage model, rather than the second stage model (at which point the assumptions are mis-specified in the latter situation). Although in a simulation study it is immediately obvious whether variables are associated with treatment choice and/or outcome, in a real life situation this is not so straightforward, and may be unverifiable. This is a limitation of the sample selection approach.

There were a number of situations in which a wide range of treatment estimates were obtained. Firstly, when I assumed that there were 500 individuals included in the study; a much smaller sample than used in the other simulations, wide ranges of treatment estimates were obtained. Stolzenberg and Ralles [406] and Sartori [407] have both similarly commented on the limitations of using sample selection models when there is a small sample size. A wide range of treatment estimates were also obtained when there were no factors associated with, or there were factors only weakly associated with, treatment

allocation. A wide range of estimates of the treatment effect were obtained, even though there were known confounders that were also associated with treatment allocation included in the models. In real life situations, it will not always be possible to identify a known independent variable only associated with treatment allocation, and again this could limit the use of sample selection models practice. Others have also commented on the wide range of values obtained from models, and this is clearly a limitation of the approach [405;407].

When considering estimates of the association between the IMR and response, it is clear that in a number of situations there was great variability in the range of estimates obtained. This was particularly apparent in situations in which the binary confounder was assumed to be unknown, and the associated p-value was not significant at the 5% level on far more occasions than one would wish. Therefore, when applying sample selection models, it appears that it is important to consider the impact of including the IMR in the second stage model on the estimate of the treatment effect, rather than purely looking for statistical significance.

Sample selection models are employed under a number of assumptions that have not been investigated here, but which have been identified previously. Models are highly sensitive to the assumption of bivariate normality, and it is also assumed that the error terms have a constant variance [408]. Heckman [397] also acknowledges that the standard errors may be downwards biased (underestimated), a limitation that has been reiterated by other authors [405]. Furthermore, it is assumed that the treatment effect is constant for all individuals participating in the study; an assumption that I felt it was reasonable to assume, but another assumption nonetheless. Finally, Stolzenberg and Rolles highlight the fact in their data simulations that seriously inflated standard errors as a result of colinearity between the IMR term and the regressors can occur [406].

The results of my data simulations also highlighted the impact of unmeasured confounding on the results of standard logistic regression models and propensity score models. Although it is well known that these models are based on the assumption that all potential confounders are known and measured, it is nonetheless important to remind ourselves of the potential impact of this bias on the results when performing analyses and reading the results of others' work. It also highlights why RCTs are the gold standard approach to compare treatment regimens.

I considered a strong associations between treatment allocation and outcome with the independent variables, usually considering a log odds ratio of 1.0 (OR equal to 2.7) or

2.0 (OR of 7.4). In real life situations, the strength of associations are often not of this order of magnitude. I also only considered just one continuous and one binary confounding factor, and thought of these as the composite impact of a number of confounding factors. However, the results may have been affected if a number of confounders were present that were acting in opposite directions.

Other methods to account for unmeasured confounding have also been proposed. Rosenbaum has developed a technique to investigate the sensitivity of observational studies to hidden bias [409]. Others have developed semi and non parametric extensions of Heckman's estimator [410-412]. However, Winship and Mare [408] state that, although these methods make fewer assumptions, they typically have large standard errors and are therefore very imprecise.

I considered a simple example to investigate the use of Sample Selection models. However, the results of this example are not simple to interpret. The confidence intervals accompanying the estimates of both the treatment effect and the IMR were extremely wide. The treatment estimate obtained from the sample selection models was of a similar magnitude to that seen in the standard logistic regression model, but it is hard to know whether to conclude that the results from the standard logistic regression model should be used or not. Other examples investigated on the Royal Free Cohort have similarly found imprecise results [413]. Bushway et al suggest only including terms in the final models which are associated with the response [405]. One suggestion they give is to use a backwards selection approach with a cut-off of p=0.05. However, in the example presented in this chapter, this did not lead to more precise results. The most likely explanation for this is the high correlation between the IMR and the outcome. Bushway et al [405] describe the problems present when high co-linearity exists, and state that it is quite common when these models are applied. They suggest that in severe cases, it may be best to simply acknowledge the thread of selection bias and continue with more simple approaches. Thus, although these models may theoretically be an attractive option to account for unmeasured confounding, application in real-life settings such as the one I have considered here may not be so straightforward.

Sample selection models have been applied in a number of medical settings, and thus the issues regarding imprecise estimates found here clearly do not always apply [367;395;399-402]. There are a number of specific and general issues when considering observational datasets of HIV positive individuals that may limit the use of sample selection models in this setting. Firstly, any comparisons performed on this dataset

consider a relatively small sample size. The results of my data simulations suggest that large sample sizes are required to perform sample selection models otherwise a wide range of treatment estimates are obtained. Although there are cohorts and cohort collaborations with a larger number of individuals than are included in the Royal Free cohort, it is likely that they are still not of a magnitude to be able to carry out these methods successfully. A small number of large cohort collaborations exist, but there are very few [118;166].

I found that it was not immediately obvious which of the independent variables considered in my example were confounders, which were associated with treatment choice only, and which were only associated with outcome. The results of my data simulations highlighted the issues surrounding mis-specification of variables. Unfortunately, it is likely that there are few factors associated only with treatment choice. In my example, the only variable considered to be associated with treatment choice only was pre-HAART viral load, as I felt there was sufficient evidence from previous studies that all other factors could be potential confounders. However, there was evidence from the standard logistic regression model that this was a possible confounder. As I made the decision on which variables to include in which model *a priori*, I did not want to change the model after these results were known, but this could potentially have led to model mis-specification. Furthermore, the results of the first stage model suggest that pre-HAART viral load is not a very strong predictor of treatment allocation. Therefore, I have only considered one variable as a predictor of treatment allocation, and it may potentially be a weak predictor. Thus, in my example, the data are comparable to that investigated in situations 27 and 28 in my data simulations. I found that in these situations that the usefulness of sample selection models was limited. As the Royal Free cohort collects most variables that are routinely measured in observational cohorts of HIV positive individuals [414], it is likely that similar issues will be encountered if using sample selection models in other cohorts.

7.6 Summary

This chapter has considered a method to account for bias caused by unmeasured confounding when comparing the incidence of antiretroviral-related toxicities amongst those receiving different antiretroviral regimens. Sample selection models have been widely used in other settings. My data simulations suggest that these models give unbiased estimates of treatment effects in certain situations. However, specific

situations considered in the data simulations, along with the application of the methods to a real life example suggest that these methods may not always be applicable in the HIV setting. Therefore, the use of these models may be limited to sensitivity analyses, and may not always provide useful results.

Chapter 8 – Assessment of linear and non-linear associations: the impact of exposure to antiretroviral therapy on total cholesterol levels in HIV-positive children

8.1 Introduction

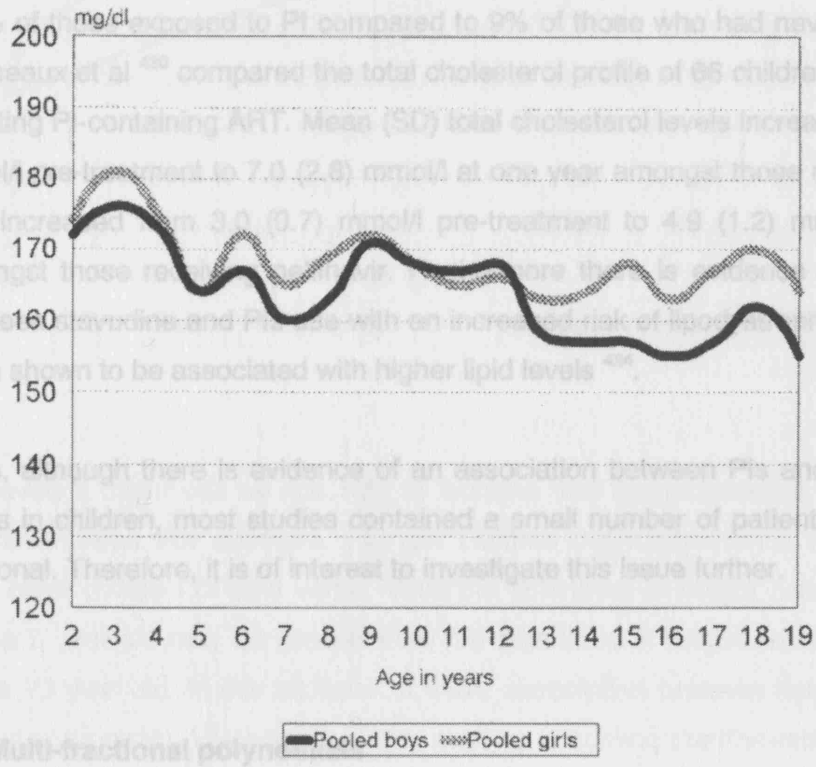
The previous chapters of this thesis have considered the potential biases present when considering the incidence and prevalence of antiretroviral-related toxicities. This chapter considers an additional problem when assessing relationships between variables, that of non-linear associations. One method proposed to investigate this issue is fractional polynomials. I shall begin this chapter by briefly describing the theory behind fractional polynomials, and shall then apply the methods to an example investigating the impact of antiretroviral therapy on total cholesterol levels in HIV-positive children. As the literature review in Chapter 2 concentrated on antiretroviral-related toxicities in adults, I shall first review the literature of the impact of antiretroviral therapy in HIV positive children on cardiovascular disease.

8.2 Literature review – antiretroviral-related metabolic disorders in HIV positive children

The risk of cardiovascular disease in the general, HIV negative, adult population has been well studied, and risk factors associated with cardiovascular disease have been established [363;415]. However, the absolute risk of cardiovascular disease is very low in children. It is estimated that between 500 and 1000 children under the age of 21 experience sudden cardiac death each year in the United States of America [416]. Therefore, the impact of changes in cardiovascular risk factors, such as blood pressure and total cholesterol, on the risk of cardiovascular disease in these children may be difficult to investigate due to the small number of events and, therefore, limited power. Nonetheless, continued exposure to risk factors in childhood could impact on long-term outcomes. Studies investigating the impact of obesity [417;418], raised cholesterol [418;419], and high blood pressure [419;420] in childhood have found an increased risk of cardiovascular disease in later life in the general HIV-negative population. Therefore, it is important to establish whether any factors, such as ART exposure, could lead to increases in proven cardiovascular risk factors in these children, and as a result lead to an increased risk of cardiovascular disease in later life.

Few studies have described the normal range of total cholesterol levels in children. The Royal Free Hospital biochemistry department quotes a normal total cholesterol range as between 1.6 and 4.9 mmol/l in children aged less than one year, between 2.8 and 6.0 mmol/l in those aged one to 18 years, and less than 5.2 mmol/l thereafter. Figure 8.1 shows the estimated total cholesterol patterns from a review study by Brotons et al of more than 60,000 children aged 2 to 19 years from 26 countries [421]. Although there may be questions surrounding the biological plausibility of this model, there appears to be a reduction in total cholesterol levels until around the age of puberty at which point they appear to stabilise. A similar pattern of decreasing total cholesterol levels after infancy, followed by an increase in early puberty and subsequent decrease until the age of 15 years was found by Boulton et al in a cohort of 586 children [422].

Figure 8.1 – Predicted distribution of total cholesterol according to age (from Brotons et al [421])



There have been a few studies investigating the prevalence of hypercholesterolaemia in HIV-positive children [423-427]. Amaya et al [423] studied forty children receiving ART aged 2 to 16 years, and found that 27 (68%) exhibited evidence of hypercholesterolaemia (>171 mg/dl [4.5 mmol/l]). The European Lipodystrophy Group [424] collected metabolic marker data on 280 children. Hypercholesterolaemia (>200 mg/dl) was observed in 25% (95% CI 21.6%, 32.7%) of children. Finally, Jaquet et al studied 39 HIV-positive

children aged >3 years cross-sectionally, and found comparable mean (SD) total cholesterol levels amongst those with (4.85 [0.89] mmol/l) and without (4.47 [0.8] mmol/l) lipodystrophy [425].

There have also been a number of studies that have investigated the impact of antiretroviral therapy, particularly PI use, on total cholesterol levels [423;424;426;428-435]. However, most of these studies are small and of a cross-sectional nature. A recent longitudinal study by Rhoads et al of 146 children from St Mary's Hospital [428], who are included in this present analysis, found that after two years of ART of any drug class total cholesterol levels had increased by 0.93 mmol/l ($p<0.0001$) compared to pre-treatment values; levels reached a plateau at this time point. Lainka et al [429] studied 37 children aged 1 to 17 years receiving ART with or without a PI. They found that the PI group had higher total cholesterol levels (mean [SD] 235 [71] mg/dl) compared to the non-PI group (176 [25] mg/dl). The European Lipodystrophy Group [424] found that hypercholesterolaemia was more common amongst those who had ever received a PI (37% of those exposed to PI compared to 9% of those who had never received a PI). Cheseaux et al [430] compared the total cholesterol profile of 66 children before and after initiating PI-containing ART. Mean (SD) total cholesterol levels increased from 3.3 (0.7) mmol/l pre-treatment to 7.0 (2.8) mmol/l at one year amongst those receiving ritonavir, and increased from 3.0 (0.7) mmol/l pre-treatment to 4.9 (1.2) mmol/l at one year amongst those receiving nelfinavir. Furthermore there is evidence of an association between stavudine and PIs use with an increased risk of lipodystrophy, which has also been shown to be associated with higher lipid levels [424].

Thus, although there is evidence of an association between PIs and total cholesterol levels in children, most studies contained a small number of patients and were cross sectional. Therefore, it is of interest to investigate this issue further.

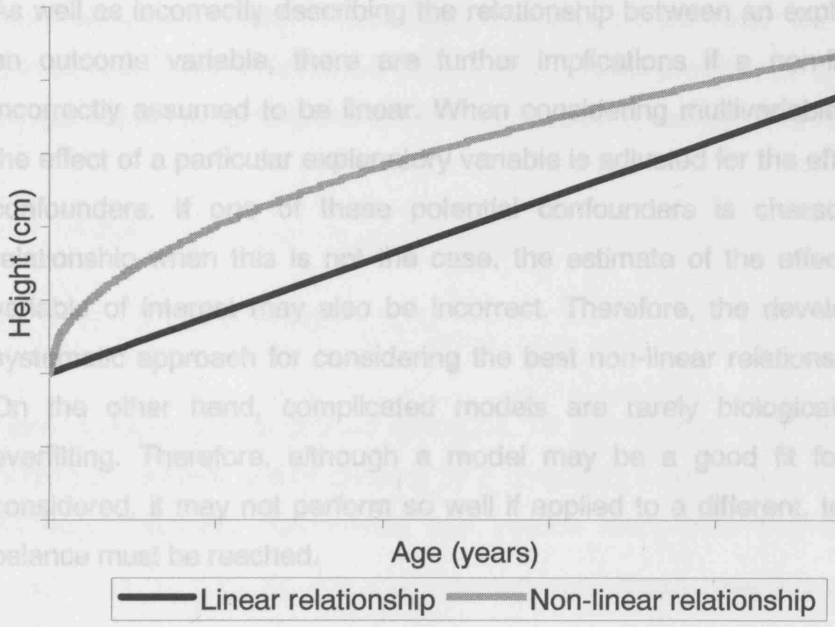
8.3 Multi-fractional polynomials

8.3.1 Non-linear associations

As described in Appendix B, linear regression models are often used to investigate factors associated with continuous outcomes, such as the relationship between age and total cholesterol levels. We usually assume that linear relationships exist between variables.

For example, consider a model in which we believe that the age of a child is associated with their height. For now we will assume that no other factors are associated with a child's height. We could start by assuming that there was a linear relationship between the two variables. Thus, for each year increment in age, a child is expected to be *x* cm taller (Figure 8.2).

Figure 8.2 – Examples of a linear and of a non-linear (square root) relationship



However, it might not be accurate to assume that the association between age and height is linear. For example, younger children might experience more rapid growth than older children or vice versa. Thus the difference in height between a 6 year old and a 7 year old may be greater than the difference in height between a 12 year old and a 13 year old. In this situation, a linear association between height and age might not be appropriate. We could perhaps start by assuming that the relationship was on a square-root scale, as shown in Figure 8.2. Thus, the increase in height is greatest at younger ages. We can fit this relationship easily in a linear regression model by taking the square root of each child's age, and fitting this as an explanatory factor instead of the child's age.

It is possible to use other non-linear functions to describe the relationship between age and height. For example, it is possible to use a polynomial of any order ($...x^{-3}$, x^{-2}, x^{-1}, x, x^2, x^3,...) as well as logarithmic functions (e.g. $\log_e(x)$, $\log_2(x)$, $\log_{10}(x)$, e^x, 2^x, 10^x)

and any combination of them. Thus, we are left with an infinite number of potential relationships that could be considered between our two variables. Although this enables great flexibility in constructing models, it also means that it is hard to decide on the most appropriate non-linear association between two variables, particularly in multivariable analysis when a there may be a number of variables to consider. On the other hand, if too simple a non-linear relationship were chosen, for example a quadratic relationship, this may not be adaptable enough to capture the relationship between the variables, particularly at the extremities.

As well as incorrectly describing the relationship between an explanatory variable and an outcome variable, there are further implications if a non-linear relationship is incorrectly assumed to be linear. When considering multivariable regression models, the effect of a particular explanatory variable is adjusted for the effect of other potential confounders. If one of these potential confounders is characterised as a linear relationship when this is not the case, the estimate of the effect of the explanatory variable of interest may also be incorrect. Therefore, the development of a simple, systematic approach for considering the best non-linear relationship would be useful. On the other hand, complicated models are rarely biologically plausible due to overfitting. Therefore, although a model may be a good fit for the dataset being considered, it may not perform so well if applied to a different, test, dataset. Thus, a balance must be reached.

8.3.2 Fractional polynomials

The methodology for multivariable fractional polynomials was developed by Royston and Altman in 1994 [436]. These models build on the idea of choosing the most appropriate non-linear relationship between variables in standard regression models in a systematic way. The authors developed the *mfp* procedure for the statistical software package Stata for these models (http://www.stata.com/help.cgi?mfp). Thus, practically, applying these models is no more complicated than fitting a standard regression model. I shall now explain briefly the rationale behind these models; for more details see Royston and Altman [436].

When carrying out a standard regression model with n explanatory factors x_1, x_2, x_3, ..., x_n, the model is of the form (see Appendix C):

$$g(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \varepsilon$$

where
$g(.)$ is link function
ε is error term with 0 mean

This model above assumes that all continuous variables have a linear relationship with the link function, *g*. Once the restriction of a linear relationship is removed, each explanatory factor, x_i, can now take any form (i.e. the term $\beta_i x_i$ is replaced with a more complex formula). When using fractional polynomials, we consider a family of curves whose power terms are restricted to a small pre-defined set of integer and non-integer values by adding the restriction that each x_i, can take the following form:

$$\alpha_i x_i^{\,p} + \gamma_i x_i^{\,q}$$

$$(p,q) \in \{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$$

Here, a power of 0 is defined as $\log_e(x)$. All variables can be linearly transformed (by adding a constant and/or dividing by a constant) if required to ensure model stability. Although fractional polynomials can include more than two terms if desired (i.e. $\alpha_i x_i^{\,p} + \gamma_i x_i^{\,q} + \lambda_i x_i^{\,r} + ... + \omega_i x_i^{\,z}$), Royston and Altman state that two terms are usually sufficient to capture most relationships. Furthermore, simpler relationships containing only one term (i.e. of the form $\alpha_i x_i^{\,p}$) can also be constructed if thought to be appropriate.

An additional statistical feature (see Royston and Altman for details [436]) of this type of model is that relationships of the form $\alpha_i x_i^{\,p} + \gamma_i (\log_e(x_i) * x_i^{\,p})$ are also permitted. Furthermore, although in the models I have applied in this chapter consider only the powers in the set $\{-2, -1, -0.5, 0, 0.5, 1, 2, 3\}$ it is also possible to extend this set. Again, this restricted set is usually sufficient for most applications.

The most appropriate form for each continuous explanatory factor can be chosen in a number of ways. For example, nested models can be compared using the change in deviance which will asymptotically follow a chi-squared distribution and allow models to be compared and p-values to be calculated. Models can also be compared heuristically by drawing figures and considering the biological plausibility of models. In my models, I

have initially selected the most appropriate form for each continuous explanatory factor using the Royston and Altman model-selection algorithm [436], with the added restriction that they must contain two terms. I have then decided whether to include each covariate as a non-linear term or not by considering plots of the data and the p-values associated with the model-selection algorithm when comparing the model to one with a simple linear term included.

In this way, there is a flexible choice of non-linear relationships that can be considered systematically and applied simply in real-life situations. I shall now go on to illustrate fractional polynomials further by considering a real-life example, the association between total cholesterol levels and antiretroviral therapy in HIV-positive children.

8.4 The impact of antiretroviral therapy on total cholesterol levels in HIV positive children - Methods

Children included in this study are from Great Ormond Street Hospital, London, UK and St Mary's Hospital, London, UK, with additional data from the UK CHIPS database. Data collection is described in detail in Chapter 3, and all children with at least one total cholesterol measurement from these hospitals were included.

I initially considered the first total cholesterol measurement for each child, and thus took a cross-sectional approach to the analysis. I carried out a linear regression model in which the outcome was total cholesterol level. Potential explanatory variables were age, age-adjusted height z-score, age-adjusted weight z-score, CD4 percentage, total length of PI use, total length of non-PI ART use, gender, ethnicity and prior exposure to mono/dual therapy. As it is unknown whether total cholesterol levels would increase linearly with age and length of exposure to ART, I carried out a multivariable fractional polynomial model. All continuous variables were investigated to look for evidence of non-linearity. I next carried out a standard linear regression model, but fitting any continuous variables with any evidence of non-linearity in the form suggested by the fractional polynomial model. I then fitted a second model fitting all variables assuming a linear relationship with total cholesterol to investigate the impact of the non-linear relationship on the estimates of other variables with total cholesterol. In a multivariable analysis, all factors with $p<0.1$ were initially included in a model (except for total ART exposure, as this is already captured by the inclusion of PI and non-PI exposure) and backward selection was used to find a final model using an exclusion criterion of $p>0.05$.

I then considered all total cholesterol measurements for each child, rather than simply the first measurement. I first performed exploratory analyses by dividing children into two-year age groups. I calculated the median and 95% CI total cholesterol level for all measurements taken on children in each age group, regardless of which child the measurement was from. The confidence intervals were calculated using the binomial based method [437]. Length of exposure to ART and to PI- and non-PI containing ART was plotted similarly, by creating half-yearly intervals. I then fitted a mixed effects model to investigate the factors associated with total cholesterol levels. First of all, I fitted age in the form suggested by the multi-fractional polynomial model performed above. The model intercept and age were fitted as random effects, and all other variables were considered as fixed effects. I compared this model to one obtained after fitting the results with a linear association for all continuous variables. This meant I could investigate the impact of the non-linear relationship on the estimated parameter estimates. All variables, except for the treatment variables, were fitted both individually and as interaction terms with age, to investigate whether, for example, boys had higher initial total cholesterol levels throughout as well as whether they changed at a different rate compared to girls. The treatment variables measure cumulative exposure to treatment, and thus are already interaction terms. Again, in a multivariable analysis, all factors with $p<0.1$ were initially included in a model, (except for non-PI and PI-use as current ART use was included instead) and backward selection was used to find a final model using an exclusion criterion of $p>0.05$.

8.5 The impact of antiretroviral therapy on total cholesterol levels in HIV positive children – Results

8.5.1 Patient population

In total 385 children were included in the analyses. At the time of entry into the study (the time at which each child had their first recorded total cholesterol measurement,) 185 were receiving antiretroviral therapy (ART), and 200 were not. During the course of the study, there were a total of 5560 total cholesterol measurements (median 14 [IQR 5, 21] per child) and 1445 person-years of follow-up (median 4.0 years [IQR 2.1 years, 5.2 years] per child). The characteristics of the children are described in Table 8.1 according to treatment use at study entry.

On average, those not receiving ART had their first total cholesterol measurement at a later time point (Median April 2001 for those not receiving ART compared to October

2000 for those receiving ART; p=0.01), a lower CD4 percentage (median 22% [IQR 16%, 28%] compared to 24% [18%, 33%]; p=0.04), a higher viral load (median 25400 copies/ml [8272, 138300] compared to 1350 [79, 31800]; p<0.0001) and a higher age-adjusted height z-score (-0.3 [-1.2, +0.3] compared to –0.6 [-1.4, +0.1]; p=0.04) compared to those not receiving ART. Other variables were comparable between groups. The majority of children were of black African ethnicity (286; 74.3%), and were known to have been vertically infected with HIV (370; 96.1%). The median [IQR] range at study entry was 6.3 (2.9, 9.4) years, and approximately half were male (195; 50.7%).

Amongst those who were not currently receiving ART, 21 (11.4%) had previously done so. Amongst those currently receiving ART, the average length of exposure was 1.7 years (IQR 0.3, 3.8 years) and around a third (63; 31.5%) had received mono or dual therapy prior to receiving HAART. As can be seen in Table 8.1, the median (IQR) total cholesterol was 3.6 mmol/l (3.1, 4.2 mmol/l) for those children not receiving ART, and was 4.2 mmol/l (3.6, 4.9 mmol/l) amongst children currently receiving ART (P<0.0001). Thus, there was some evidence from an unadjusted comparison that total cholesterol levels were higher amongst those receiving ART compared to those who were not.

Table 8.1 – Characteristics of HIV-positive children from Great Ormond Street and St Mary's Hospitals at the time of their first total cholesterol measurement

		Number (%)			
		Not receiving ART	**Receiving ART**	**Total**	**p-value**
Number		185 (100.0%)	200 (100.0%)	385 (100.0%)	
Total cholesterol (mmol/l)	*Median (IQR)*	3.6 (3.1, 4.2)	4.2 (3.6, 4.9)	3.9 (3.3, 4.5)	<0.0001
Date of measurement	*Median (IQR)*	April 2001 (Jul 2000, Mar 2003)	Oct 2000 (Feb 2000, Aug 2002)	Jan 2001 (May 2000, Nov 2002)	0.01
Ethnicity	*Black African*	144 (77.8)	142 (71.0)	286 (74.3)	0.28
	White	13 (7.0)	25 (12.5)	38 (9.9)	
	Mixed	22 (11.9)	23 (11.5)	45 (11.7)	
	Other	6 (0)	10 (1.0)	16 (0.5)	
Age	*Median (IQR)*	6.2 (3.1, 8.8)	6.5 (2.7, 10.0)	6.3 (2.9, 9.4)	0.38
Vertically infected	*Yes*	179 (96.8)	191 (95.5)	370 (96.1)	0.52
Gender	*Male*	86 (46.5)	109 (54.5%)	195 (50.7%)	0.12
Absolute CD4 count (cells/mm^3)	*Median (IQR)*	591 (390, 878)	686 (344, 1143)	610 (367, 1010)	0.04
CD4 percentage	*Median (IQR)*	22 (16, 28)	24 (18, 33)	23 (17, 30)	0.03
Viral load (copies/ml)	*Median (IQR)*	25400 (8272, 138300)	1350 (79, 31800)	10600 (700, 71857)	<0.0001
Height (cm)	*Median (IQR)*	118 (98, 131)	115 (92, 132)	115 (95, 132)	0.49
Age-adjusted height z-score	*Median (IQR)*	-0.3 (-1.2, +0.3)	-0.6 (-1.4, +0.1)	-0.5 (-1.4, +0.2)	0.04
Weight (kg)	*Median (IQR)*	21.6 (15.1, 28.1)	20.4 (15.1, 29.7)	21.5 (15.1, 29.0)	0.83
Age-adjusted weight z-score	*Median (IQR)*	-0.1 (-0.9, +0.5)	-0.2 (-1.0, +0.5)	-0.1 (-1.0, +0.5)	0.35
Current ART	*PI*	-	91 (45.5)	-	-
	NNRTI	-	97 (48.5)	-	
Current antiretrovirals	*AZT*	-	82 (41)	-	-
	ddI	-	65 (32.5)	-	
	3TC	-	127 (63.5)	-	

		Not receiving ART	Number (%) Receiving ART	Total	p-value
	d4T	-	90 (45)	-	
	ABC	-	54 (27.0)	-	
	NVP	-	67 (33.5)	-	
	EFV	-	30 (15)	-	
	NFV	-	73 (36.5)	-	
Ever received ARVs	*Yes*	21 (11.4%)	200 (100.0%)	221 (57.4%)	-
Length of exposure to ARVs (years)	*Median (IQR)*	0 (0, 0)	1.7 (0.3, 3.8)	-	-
Length of exposure to PIs (years)	*Median (IQR)*	0 (0, 0)	0.0 (0.0, 1.5)	-	-
Length of exposure to NNRTIs (years)	*Median (IQR)*	0 (0, 0)	0.5 (0.02, 2.3)	-	-
Length of exposure to NRTIs (years)	*Median (IQR)*	0 (0, 0)	1.7 (0.3, 3.8)	-	-
Received mono/dual therapy prior to HAART	*Yes*	16 (8.7)	63 (31.5)	79 (20.5%)	-

IQR=inter quartile range; BMI=body mass index; ARV=antiretroviral; PI=protease inhibitor; NNRTI=non-nucleoside reverse transcriptase inhibitor; NRTI=nucleoside reverse transcriptase inhibitor; AZT=zidovudine; ddI=didanosine; 3TC=lamivudine; d4T=stavudine; ABC=abacavir; NVP=nevirapine; EFV=efavirenz; RTV=ritonavir; NFV=nelfinavir; HAART=highly active antiretroviral therapy (defined as 3 or more antiretrovirals). p-values calculated using Wilcoxon test, chi-squared test and Fisher's exact test as appropriate

8.5.2 Choice of fractional polynomial

I then went on to consider each child at the date of his or her first total cholesterol measurement (i.e. study entry) and investigated factors associated with this first total cholesterol measurement using a multivariable fractional polynomial. The model was fitted such that all of the continuous variables were required to have a non-linear relationship, regardless of the significance of the effect. The results of this model are shown in Table 8.2. Considering first the effect of age on total cholesterol levels, it can be seen from the associated p-value that there is some evidence, albeit not statistically significant at the 5% level, that fitting a non-linear relationship between age and total cholesterol leads to a better fitting model than fitting a linear relationship (p=0.083). Note that this p-value considers whether there is evidence of a non-linear relationship between variables, as opposed to whether an association between total cholesterol and age exists at all. As the sample size is relatively small (385 children), it is therefore worth investigating this relationship further. The relationship suggested by the model suggests including a squared age term, as well as a squared term multiplied by the natural log of age. The exact coefficients are shown in Table 8.2.

Table 8.2 – Factors associated with first total cholesterol measurement: multi-factorial polynomials method to investigate non-linear associations

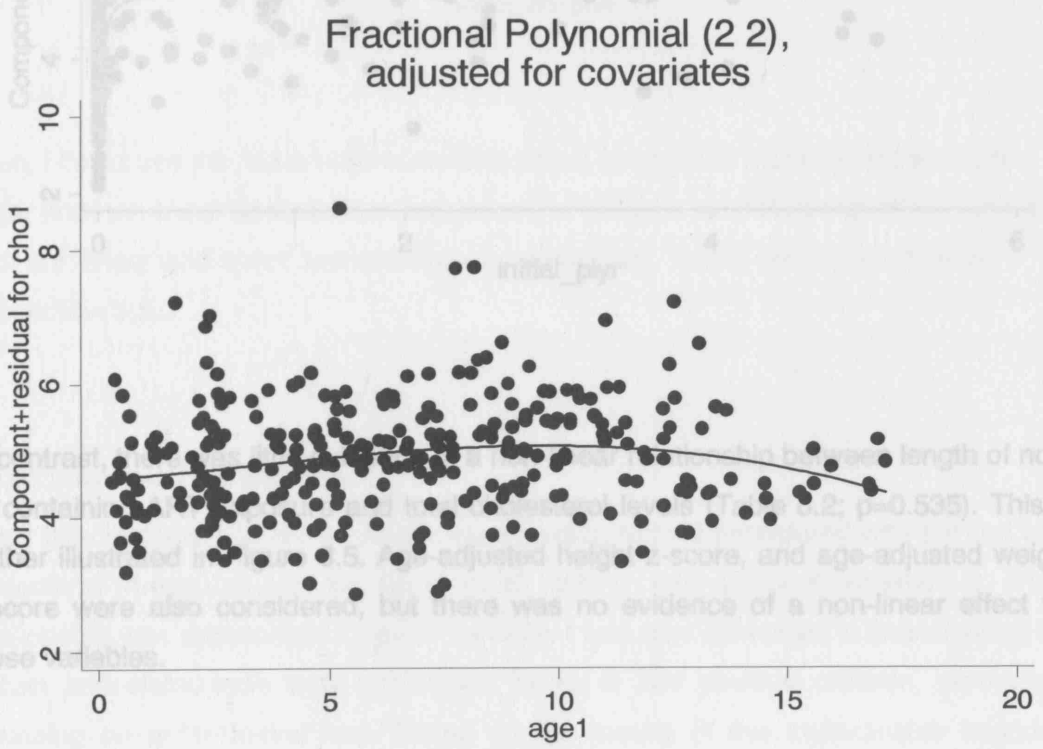
Variable	**Form of relationship between TC and variable**	X_1	X_2	**p***
Age (age)	$X_1^2 + (X_2^2 * (\log_e [X_2]))$	age^2-0.4556	$(age/10)^2$*ln((age1/10))+0.1791	0.08
CD4% (cd4p)	$(X_1)^3 + (X_2)^{-0.5}$	$(cd4p)^{-0.5}$-0.6451	$(cd4p/10)^3$-13.8698	0.64
Height z-score (htz)	$(X_1)^{-2} + (X_2)^{-0.5}$	$(htz)^{-2}$-0.5304	$(htz+14.3034)/10)^{-0.5}$ -0.8534	0.99
Weight z-score (wtz)	$(X_1)^3 + (X_2^3 * (\log_e [X_2]))$	$(wtz)^3$-28.0680	$(wtz+3.2443)^3$ *ln((wtz+3.2443))-31.1988	0.63
Exposure to PIs (epi)	$(X_1)^{-2} + (X_2)^{-1}$	$(epi)^{-2}$-3.4708	$(epi+0.0027)^{-1}$-1.8630	0.001
Exposure to non-PIs (enp)	$(X_1)^{-0.5} + (X_2)$	$(enp)^{-0.5}$-3.3794	(enp+0.0027)/10)-0.0876	0.54

Other (categorical) variables included in the model: ethnicity, gender, previous mono/dual therapy; TC=total cholesterol

Therefore, the multi fractional polynomial model has provided an estimate of the form of the relationship between age and total cholesterol. However, it is hard to visualise

what this relationship looks like. Therefore, I have illustrated this in Figure 8.3, using a residual scatter plot. Each child is represented by a point on the graph, plotting the child's residual component of total cholesterol against his or her age. Thus, this plot is adjusted for other covariates in the model and describes the relationship between the two variables estimated by the fractional polynomial model. For this line, you can see that there is a general increase in total cholesterol in the first years after birth; levels then plateau before beginning to decrease at around the age of 10 years of age. This is not contradictory with the studies mentioned in the literature review which found an increase in total cholesterol levels until the age of puberty, followed by a decrease in levels. Therefore, there is some evidence that a linear association between age and total cholesterol may not be the most appropriate to describe the relationship.

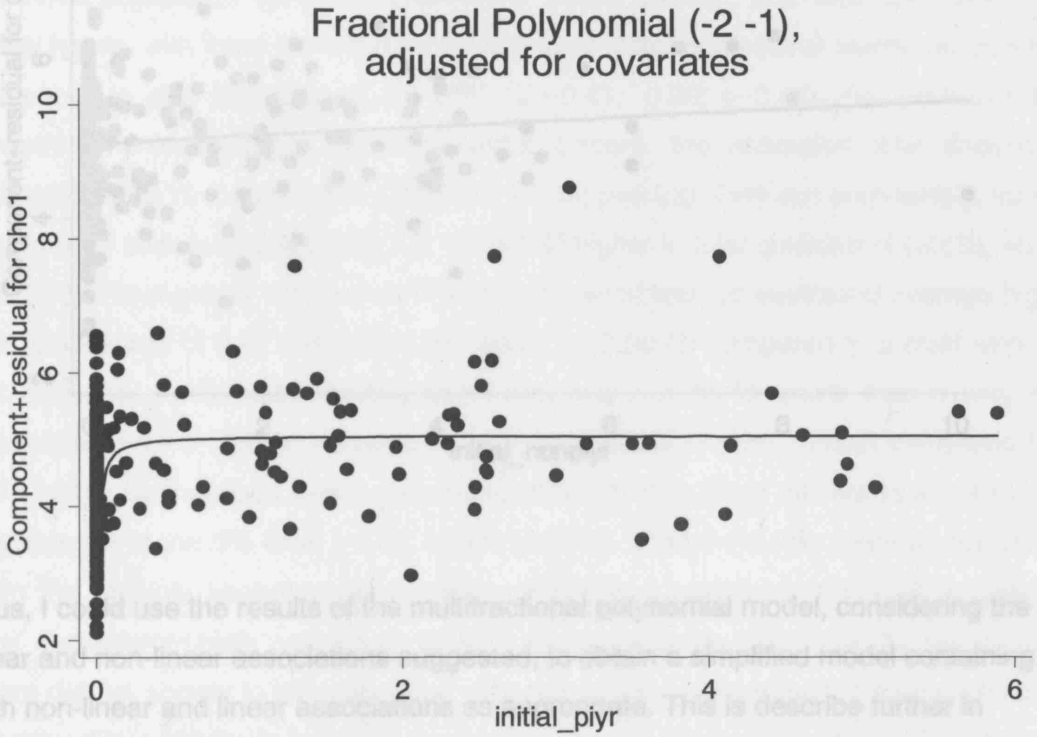
Figure 8.3- Association between total cholesterol and age predicted by multi-fractional polynomial model



There is also evidence of a non-linear relationship between length of exposure to PI-containing ART and total cholesterol levels (Table 8.2; p=0.001). This relationship consists of a squared term and a reciprocal term (length of exposure to the power of minus one). Figure 8.4 displays this relationship graphically. It can be seen that there is

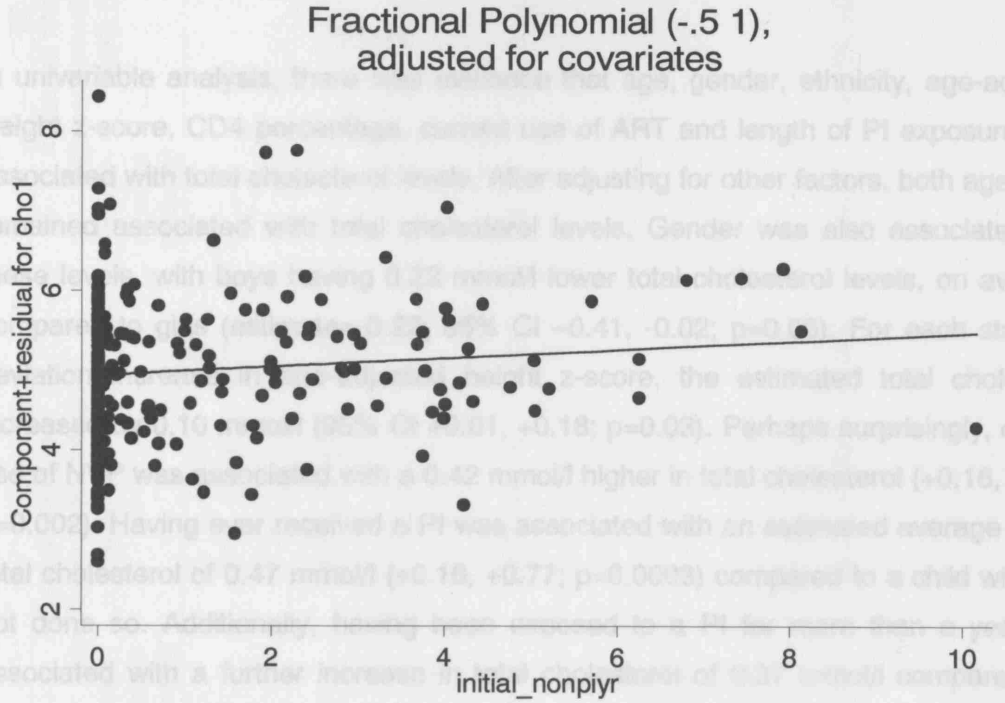
an initial steep increase in total cholesterol levels in the first few months after starting PI-containing ART, which then plateaus.

Figure 8.4 - Association between length of exposure to PI-containing ART and total cholesterol predicted by multi-fractional polynomial model



In contrast, there was little evidence of a non-linear relationship between length of non-PI containing ART exposure and total cholesterol levels (Table 8.2; p=0.535). This is further illustrated in Figure 8.5. Age-adjusted height z-score, and age-adjusted weight z-score were also considered, but there was no evidence of a non-linear effect for these variables.

Figure 8.5 - Association between length of exposure to non-PI-containing ART and total cholesterol predicted by multi-fractional polynomial model



Thus, I could use the results of the multifractional polynomial model, considering the linear and non-linear associations suggested, to obtain a simplified model containing both non-linear and linear associations as appropriate. This is describe further in Subsection 8.5.3

8.5.3 Cross-sectional analysis

I next fitted a linear regression model investigating factors associated with the first total cholesterol measurement for each child, taking into account whether a linear relationship was appropriate for each variable. I was now interested in investigating the factors associated with total cholesterol levels in HIV positive children, particularly focussing on antiretroviral use. Based on the results of the multivariable fractional polynomial model, I chose to fit age as a non-linear relationship in the way suggested (i.e. age squared and (age squared)*$\log_e x$). As there was evidence of a non-linear association between PI-containing ART exposure and total cholesterol levels, I decided to fit all of the ART variables considered as being non-linear. However, I decided to fit all ART variables as categorical variables, rather than as non-linear continuous variables (although length of exposure to ART was not fitted in the same model as

length of exposure to PI- and non PI-containing ART). The reason for this is that I wished to be able to investigate the time at which the effect of ART exposure started to plateau, and also so that I could obtain easily interpretable estimates.

In univariable analysis, there was evidence that age, gender, ethnicity, age-adjusted weight z-score, CD4 percentage, current use of ART and length of PI exposure were associated with total cholesterol levels. After adjusting for other factors, both age terms remained associated with total cholesterol levels. Gender was also associated with these levels, with boys having 0.22 mmol/l lower total cholesterol levels, on average, compared to girls (estimate=-0.22; 95% CI –0.41, -0.02; p=0.03). For each standard deviation increase in age-adjusted height z-score, the estimated total cholesterol increased by 0.10 mmol/l (95% CI +0.01, +0.18; p=0.03). Perhaps surprisingly, current use of NVP was associated with a 0.42 mmol/l higher in total cholesterol (+0.16, +0.68; p=0.002). Having ever received a PI was associated with an estimated average higher total cholesterol of 0.47 mmol/l (+0.16, +0.77; p=0.0003) compared to a child who had not done so. Additionally, having been exposed to a PI for more than a year was associated with a further increase in total cholesterol of 0.37 mmol/l compared to a child with less than one year's exposure, although this effect did not reach statistical significance at the 5% level (-0.07, +0.81; p=0.10). Therefore, this confirms our finding from the multiple fractional polynomial model that a non-linear association between total cholesterol levels and length of ART exposure exists. Exposure for two or more years did not appear to be associated with any additional increases in total cholesterol; compared to a child with between one and two year's exposure to PI, a child with more than 3 years' exposure had, on average, a total cholesterol level that was 0.03 mmol/l higher (-0.50, +0.56; p=0.92).

Table 8.3 - Factors associated with initial total cholesterol measurement; firstly assuming a non-linear relationship between age and total cholesterol, and secondly assuming a linear relationship between age and total cholesterol

		Unadjusted analysis			Adjusted analysis (non linear association)			Adjusted analysis (linear association)		
		Mean change (mmol/l)	**95% CI**	**p-value**	**Mean change (mmol/l)**	**95% CI**	**p-value**	**Mean change (mmol/l)**	**95% CI**	**p-value**
Age (years)	*Squared*	+0.47	+0.19, +0.75	0.001	+0.46	+0.17, +0.76	0.002			
	*Squared*ln(age)*	-0.91	-1.52, -0.30	0.003	-1.00	-1.61, -0.39	0.001			
Age (years)	*Per year older*	+0.03	+0.00, +0.05	0.03						
Gender	*Male vs Female*	-0.18	-0.38, +0.01	0.07	-0.22	-0.41, -0.02	0.03	-0.22	-0.41, -0.02	0.03
Ethnicity	*Black African vs. other*	-0.21	-0.44, +0.01	0.06						
Height z-score	*Per 1 S.D. higher*	+0.05	-0.03, +0.12	0.22						
Weight z-score	*Per 1 S.D. higher*	+0.11	+0.03, +0.20	0.01	+0.10	+0.01, +0.18	0.03	+0.10	+0.02, +0.18	0.02
CD4 percentage	*Per 1% higher*	+0.02	+0.01, +0.03	0.0002						
Prior mono/dual therapy	*Yes vs No*	+0.14	-0.10, +0.38	0.26						
Current use of ART	*Yes*	+0.56	+0.37, +0.75	<0.0001						
Current use of PIs	*Yes*	+0.67	+0.45, +0.90	<0.0001						
Current use of NNRTIs	*Yes*	+0.32	+0.10, +0.55	0.005						
Current use of NFV	*Yes*	+0.66	+0.42, +0.91	<0.0001						
Current use of NVP	*Yes*	+0.28	+0.02, +0.54	0.03	+0.42	+0.16, +0.68	0.002	+0.34	+0.08, +0.60	0.01
Current use of EFV	*Yes*	+0.28	-0.09, +0.65	0.14						
Ever exposed to ART	*Yes*	+0.34	+0.09, +0.59	0.008						
Effect of additional ART exposure	*>1 year*	+0.44	+0.07, +0.81	0.02						
	>2 years	-0.02	-0.48, +0.44	0.94						
	>3 years	-0.24	-0.76, +0.28	0.36						

		Unadjusted analysis			Adjusted analysis (non linear association)			Adjusted analysis (linear association)		
		Mean change (mmol/l)	**95% CI**	**p-value**	**Mean change (mmol/l)**	**95% CI**	**p-value**	**Mean change (mmol/l)**	**95% CI**	**p-value**
	>4 years	+0.19	-0.28, +0.66	0.43						
Ever exposed to PIs	*Yes*	+0.31	+0.00, +0.62	0.05	+0.47	+0.16, +0.77	0.003	+0.39	+0.08, +0.69	0.01
Effect of additional PI exposure	*>1 year*	+0.56	+0.11, +1.02	0.02	+0.37	-0.07, +0.81	0.10	+0.44	+0.00, +0.88	0.05
	>2 years	+0.09	-0.47, +0.64	0.76	+0.03	-0.50, +0.56	0.92	+0.09	-0.45, +0.62	0.75
	>3 years	-0.33	-1.08, +0.42	0.39	-0.12	-0.93, +0.68	0.76	-0.06	-0.88, +0.76	0.88
	>4 years	+0.10	-0.70, +0.90	0.81	+0.07	-0.78, +0.92	0.87	-0.03	-0.89, +0.83	0.95
Ever exposed to non-PI ART	*Yes*	+0.10	-0.15, +0.35	0.44						
Effect of additional non-PI exposure	*>1 year*	+0.22	-0.20, +0.63	0.31						
	>2 years	+0.15	-0.38, +0.69	0.57						
	>3 years	-0.38	-0.98, +0.23	0.22						
	>4 years	+0.29	-0.31, +0.90	0.34						

95% CI=95% confidence interval; S.D=standard deviation; ART=antiretroviral treatment; PI=protease inhibitor; NNRTI=non-nucleoside reverse transcriptase inhibitor; NVP=nevirapine; EFV=efavirenz; NFV=nelfinavir. Estimates from a linear regression model.

I then went on to repeat this analysis, but this time I fitted a linear relationship between age and total cholesterol levels (Table 8.3). As expected, the unadjusted results are identical to those in the non-linear model, with the exception of age. Here, there was a very small increase, albeit statistically significant, of 0.03 mmol/l (+0.0, +0.05; p=0.03) in total cholesterol for each year older in age. When considering the results of the multivariable analysis, age was no longer selected for the final model. When considering the other covariates, the same variables were chosen as in Table 8.3. In this model boys were estimated to have total cholesterol levels that were 0.22 mmol/l lower, on average, than girls (95% CI –0.41, -0.02; p=0.03), an estimate that is virtually identical to that obtained when a non-linear age term was included (Table 8.3). Similarly, a virtually identical relationship was obtained for age-adjusted weight z-score. When considering a linear age term, the effect of current use of NVP was 0.34 (95% CI 0.08, 0.60; p=0.01), which is also similar to the estimate of +0.42 mmol/l obtained in Table 8.3. The estimates of the relationship with length of PI exposure also changed slightly. The effect of ever being exposed to PIs was an increase in total cholesterol of 0.39 mmol/l (+0.08, +0.69; p=0.01), compared to the +0.47 mmol/l observed when a non-linear age effect was considered. The additional effect of more than one year's PI exposure was now just significant at the 5% level (p=0.05), although the magnitude of the effect size did not change substantially. As before, when considering a linear age term, there was no evidence that longer exposure to PIs was associated with an additional increase in total cholesterol levels.

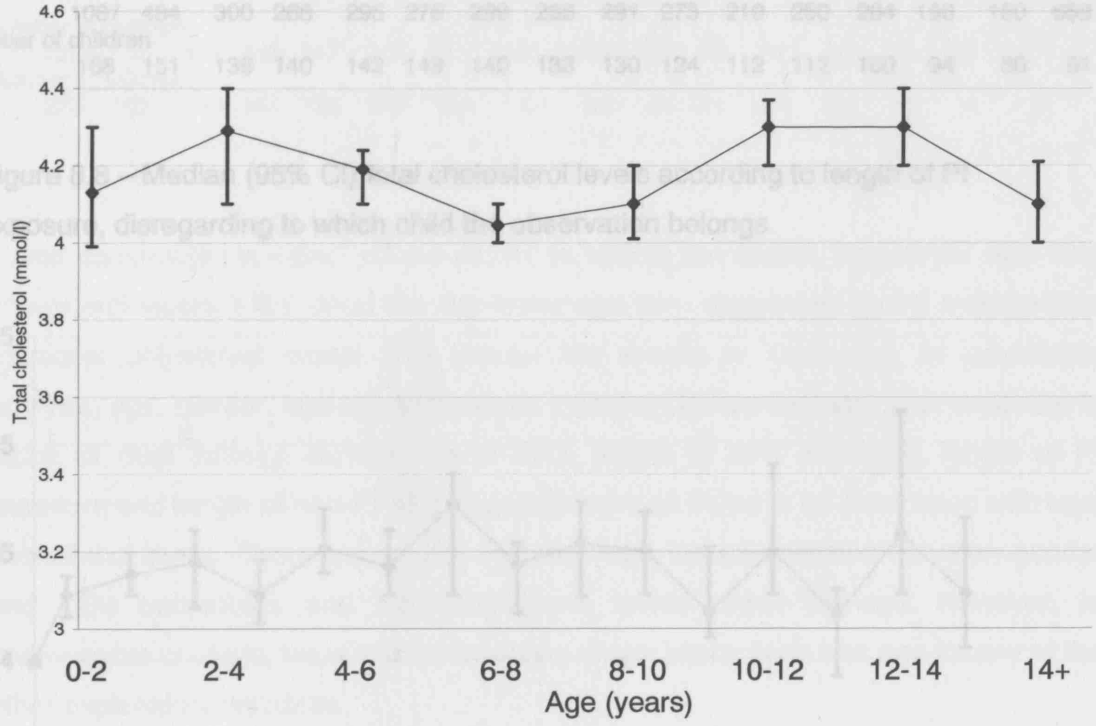
Thus, if we were not interested in the relationship between age and total cholesterol levels itself, but just wanted to obtain estimates of the relationship between ART exposure and total cholesterol levels, adjusted for other factors including age, in this particular example it is probably reasonable to present results from either regression model.

8.5.4 Longitudinal analysis

Although the results from the above models provide an insight into the association between total cholesterol levels and ART exposure, I wanted to consider all 5560 total cholesterol measurements that had been performed on the 385 children. I first carried out some exploratory analyses to investigate the relationship between total cholesterol and the continuous variables of interest. Figure 8.6 illustrates the relationship between age and total cholesterol levels, treating each observation as independent and ignoring information on the child to which observation belonged. There does not appear to be a pattern in these measurements. Figure 8.7 considers the association between length of

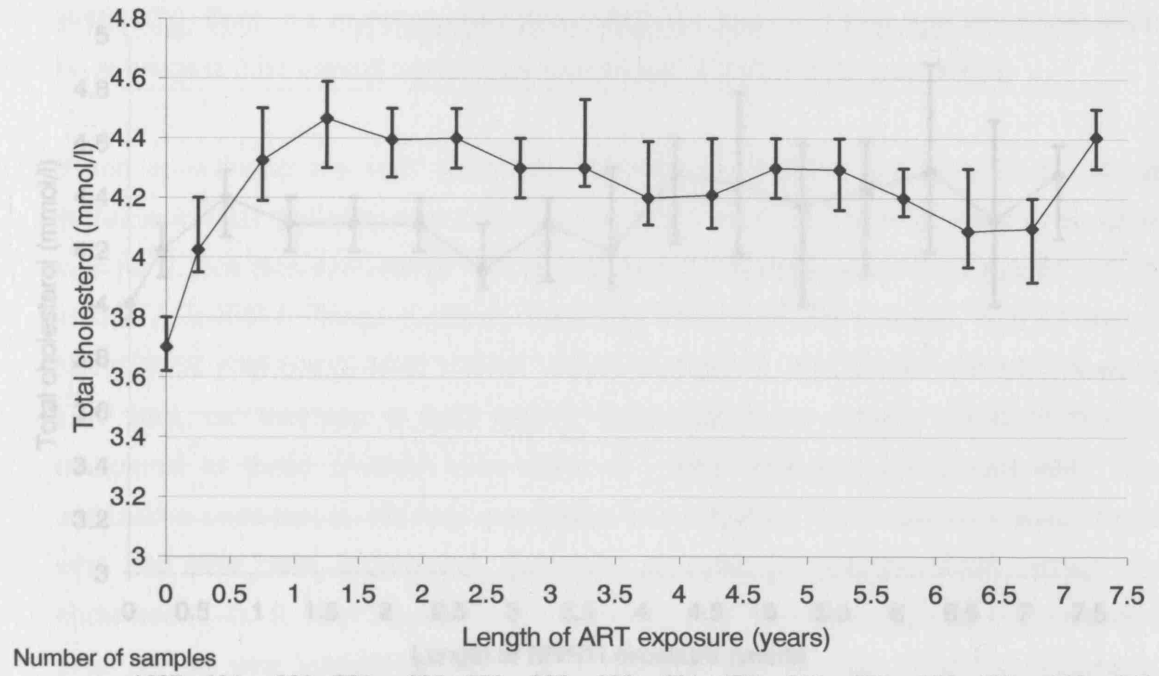
exposure to ART and total cholesterol levels. As was found in our models when considering the first total cholesterol measurement on each child, there appears to be a sharp increase in total cholesterol levels in the first one to two years of ART, before a plateau is reached. Figure 8.8 repeats this analysis, but only considers exposure to PI-containing ART. Again, there appears to be an increase in total cholesterol levels in children in the first one to two years of ART exposure, followed by a stabilising in these levels. In contrast to the results from the simple linear regression models (Table 8.3), there also appears to be a similar pattern emerging for length of exposure to non-PI containing ART (Figure 8.9), with an increase in total cholesterol levels in the first two years of ART; it should be noted that Figure 8.9 does not take into account to which child the observation belongs and is not adjusted for potential confounders.

Figure 8.6 – Median (95% CI) total cholesterol levels according to age, disregarding to which child the observation belongs



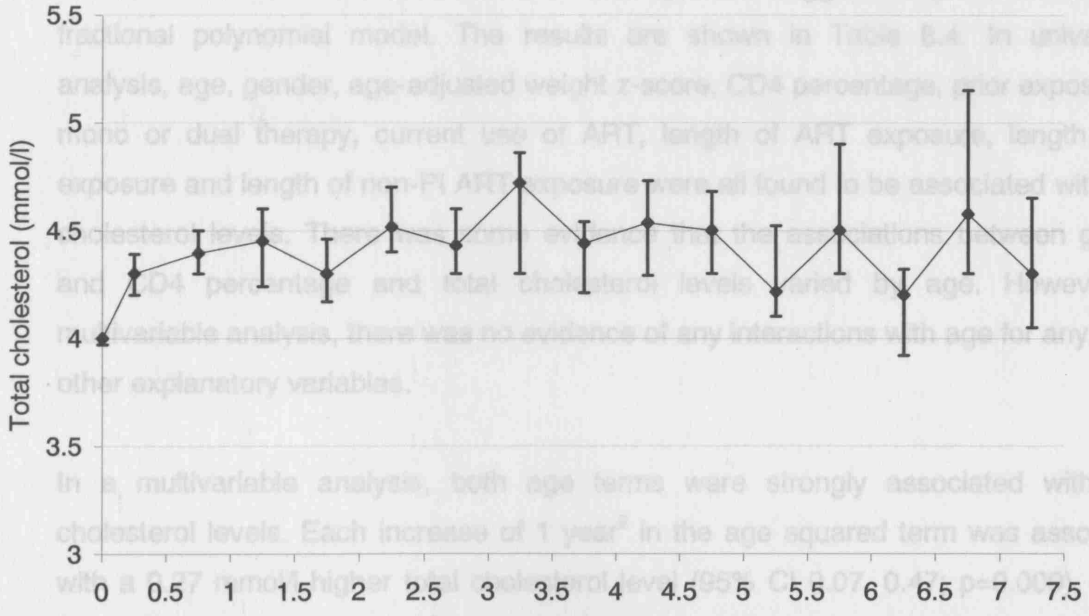
Number of samples							
315	584	817	882	973	777	656	556
Number of children							
54	116	157	167	174	158	119	94

Figure 8.7 – Median (95% CI) total cholesterol levels according to length of exposure to ART, disregarding to which child the observation belongs



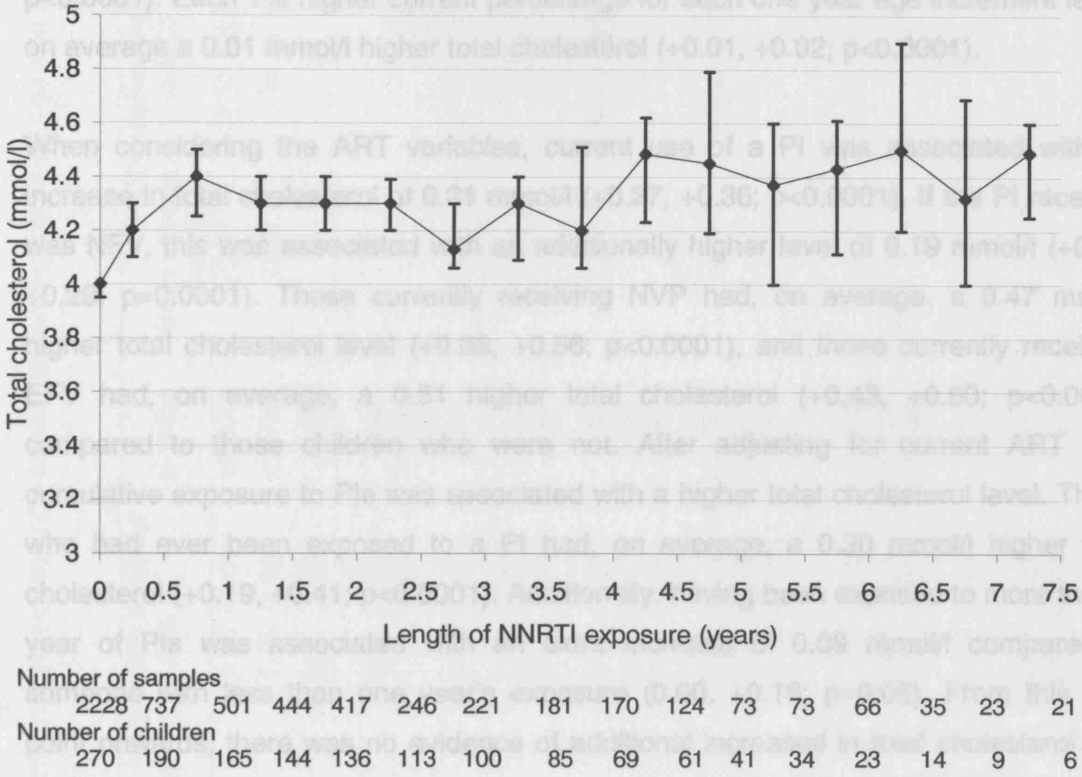
Number of samples															
1087	484	300	268	296	276	299	288	291	273	210	250	204	196	180	658
Number of children															
168	151	138	140	142	149	140	133	130	124	112	112	100	94	80	81

Figure 8.8 – Median (95% CI) total cholesterol levels according to length of PI exposure, disregarding to which child the observation belongs



Number of samples														
3008	419	238	197	237	212	235	146	166	160	149	134	71	79	52	57
Number of children															
278	90	83	89	80	80	82	76	71	65	58	44	34	26	22	17

Figure 8.9 – Median (95% CI) total cholesterol levels according to length of NNRTI exposure, disregarding to which child the observation belongs



I next constructed a mixed effects model to investigate factors associated with total cholesterol levels. I first fitted the non-linear age term suggested by the multivariable fractional polynomial model. The results are shown in Table 8.4. In univariable analysis, age, gender, age-adjusted weight z-score, CD4 percentage, prior exposure to mono or dual therapy, current use of ART, length of ART exposure, length of PI exposure and length of non-PI ART exposure were all found to be associated with total cholesterol levels. There was some evidence that the associations between gender and CD4 percentage and total cholesterol levels varied by age. However, in multivariable analysis, there was no evidence of any interactions with age for any of the other explanatory variables.

In a multivariable analysis, both age terms were strongly associated with total cholesterol levels. Each increase of 1 year^2 in the age squared term was associated with a 0.27 mmol/l higher total cholesterol level (95% CI 0.07, 0.47; p=0.009), and a one unit increase in the age^2*$\log_e$(age) term was associated with a reduction of 0.62 mmol/l in total cholesterol (-0.95, -0.29; p=0.0002). Independently of this, a one standard deviation higher height z-score combined with a one year age increment was

associated with a small decrease in total cholesterol of 0.05 mmol/l (-0.10, -0.01; p=0.01). Conversely, weight z-score was associated with a total cholesterol increase of 0.09 mmol/l per 1 standard deviation higher and 1 year age increment (+0.05, +0.14; p<0.0001). Each 1% higher current percentage for each one year age increment led to on average a 0.01 mmol/l higher total cholesterol (+0.01, +0.02; p<0.0001).

When considering the ART variables, current use of a PI was associated with an increase in total cholesterol of 0.31 mmol/l (+0.27, +0.36; p<0.0001). If the PI received was NFV, this was associated with an additionally higher level of 0.19 mmol/l (+0.10, +0.29; p=0.0001). Those currently receiving NVP had, on average, a 0.47 mmol/l higher total cholesterol level (+0.38, +0.56; p<0.0001), and those currently receiving EFV had, on average, a 0.51 higher total cholesterol (+0.43, +0.60; p<0.0001) compared to those children who were not. After adjusting for current ART use, cumulative exposure to PIs was associated with a higher total cholesterol level. Those who had ever been exposed to a PI had, on average, a 0.30 mmol/l higher total cholesterol (+0.19, +0.41; p<0.0001). Additionally, having been exposed to more than a year of PIs was associated with an extra increase of 0.09 mmol/l compared to someone with less than one year's exposure (0.00, +0.18; p=0.06). From this time point onwards, there was no evidence of additional increases in total cholesterol with increased PI exposure. Finally, there was no evidence for an association between non-PI ART exposure.

Table 8.4 – Factors associated with total cholesterol levels in HIV-positive children (assuming non-linear relationship between age and total cholesterol)

		Univariable analysis			Multivariable analysis		
		Mean change (mmol/l)	**95% CI**	**p-value**	**Mean change (mmol/l)**	**95% CI**	**p-value**
Age (years)	*Squared age*	+0.58	+0.36, +0.81	<0.0001	+0.27	+0.07, +0.47	0.009
	*Squared *ln age*	-1.03	-1.43, -0.63	<0.0001	-0.62	-0.95, -0.29	0.0002
Gender	*Male*	-0.08	-0.24, +0.08	0.35			
	Female	0.00 (ref)	-				
Age (squared)/gender interaction		+0.46	+0.02, +0.90	0.04			
Age (squared+ln)/gender interaction		-0.88	-1.66, -0.09	0.03			
Ethnicity	*African*	-0.12	-0.30, +0.06	0.19			
	Other	0.00 (ref)	-				
Age (squared)/ethnicity interaction		+0.17	-0.32, +0.66	0.50			
Age (squared+ln)/ethnicity interaction		-0.25	-1.12, +0.63	0.58			
Height z-score	*Per 1 S.D. higher*	+0.02	-0.02, +0.06	0.26	-0.05	-0.10, -0.01	0.01
Age (squared)/height interaction		-0.02	-0.12, +0.07	0.65			
Age (squared+ln)/height interaction		-0.21	-0.43, +0.01	0.06			
Weight z-score	*Per 1 S.D. higher*	+0.13	+0.09, +0.17	<0.0001	+0.09	+0.05, +0.14	<0.0001
Age (squared)/weight interaction		-0.049	-0.154, +0.057	0.37			
Age (squared+ln)/weight interaction		-0.079	-0.311, +0.153	0.51			
CD4 percentage	*Per 1% higher*	+0.02	+0.02, +0.02	<0.0001	+0.01	+0.01, +0.02	<0.0001
Age (squared)/CD4% interaction		+0.01	+0.00, +0.02	0.007			
Age (squared+ln)/CD4% interaction		-0.03	-0.05, -0.01	0.002			
Prior exposure to mono/dual therapy	*Yes*	+0.22	+0.03, +0.41	0.02			
	No	0.00 (ref)	-				
Age (squared)/monotherapy interaction		+0.24	-0.33, +0.82	0.41			
Age (squared+ln)/monotherapy interaction		-0.02	-0.99, +0.95	0.97			
Current use of ART	*Yes*	+0.69	+0.63, +0.75	<0.0001			

		Univariable analysis			Multivariable analysis		
		Mean change (mmol/l)	**95% CI**	**p-value**	**Mean change (mmol/l)**	**95% CI**	**p-value**
Current use of PIs	*Yes*	+0.52	+0.46, +0.59	<0.0001	+0.31	+0.27, +0.36	<0.0001
Current use of NNRTIs	*Yes*	+0.36	+0.30, +0.42	<0.0001			
Current use of NFV	*Yes*	+0.28	+0.19, +0.38	<0.0001	+0.19	+0.10, +0.29	0.0001
Current use of NVP	*Yes*	+0.35	+0.26, +0.43	<0.0001	+0.47	+0.38, +0.56	<0.0001
Currently receiving EFV	*Yes*	+0.34	+0.27, +0.42	<0.0001	+0.51	+0.43, +0.60	<0.0001
Ever exposed to ART	*Yes*	+0.741	+0.651, +0.830	<0.0001			
Effect of additional ART exposure	*>1 year*	+0.186	-0.114, +0.259	<0.0001			
	>2 years	+0.005	-0.072, +0.082	0.89			
	>3 years	-0.040	-0.117, +0.037	0.31			
	>4 years	+0.032	-0.043, +0.107	0.40			
Ever exposed to PIs	*Yes*	+0.63	+0.54, +0.72	<0.0001	+0.30	+0.19, +0.41	<0.0001
Effect of additional PI exposure	*>1 year*	+0.14	+0.04, +0.23	0.005	+0.09	-0.00, +0.18	0.06
	>2 years	+0.01	-0.09, +0.11	0.90	+0.01	-0.08, +0.11	0.78
	>3 years	-0.02	-0.12, +0.08	0.71	-0.04	-0.13, +0.06	0.43
	>4 years	-0.03	-0.13, +0.08	0.63	-0.04	-0.13, +0.07	0.49
Ever exposed to non-PI ART	*Yes*	+0.50	+0.41, +0.58	<0.0001	+0.04	-0.06, +0.14	0.40
Effect of additional non-PI exposure	*>1 year*	+0.13	+0.05, +0.21	0.001	+0.08	-0.01, +0.15	0.08
	>2 years	-0.05	-0.14, +0.048	0.34	-0.05	-0.14, +0.04	0.29
	>3 years	-0.13	-0.23, -0.02	0.02	-0.06	-0.16, +0.04	0.26
	>4 years	+0.08	-0.03, +0.18	0.16	+0.01	-0.10, +0.11	0.87

Using mixed effects models, with age as the underlying time variable (both parts fitted as random effects), which is included in each model

I then re-fitted this mixed effects model, but this time included a linear age effect (Table 8.5). As age was the underlying time variable and a random effect, it was included in all models regardless of its statistical significance. In unadjusted analyses, each extra year of age was associated with an increase of 0.03 mmol/l in total cholesterol (+0.01, +0.05; p=0.0005). However, after adjustment for other factors, this effect no longer remained significant (estimate=0.00; -0.02, +0.02; p=0.95), unlike when a non-linear age effect was fitted. However, the same final multivariable model was selected and, with the exceptions described below, the estimates, 95% CIs and p-values were virtually identical. When considering length of PI exposure, the model with a linear age effect found an increase of 0.27 mmol/l (+0.16, +0.38; p<0.0001) in comparison with an estimate of +0.30 (+0.19, +0.41; p<0.0001) when a non-linear age term was fitted. The additional effect of more than one year's PI exposure was +0.10 (+0.01, +0.19; p=0.03) in the linear age term model, and +0.09 (0.00, +0.18; p=0.06) when a non-linear age term was considered.

Differences between the models shown in Tables 8.4 and 8.5 were, however, evident for the current PI use and current NFV use variables. When considering a non-linear age term, as in Table 8.4, the effect of current use of a PI was estimated as an increase of 0.31 mmol/l in total cholesterol (+0.27, +0.36; p<0.0001), and the additional effect of current NFV use was estimated as +0.19 mmol/l (+0.10, +0.29). However, when considering a linear age effect (Table 8.5), the effect of current use of a PI was estimated as +0.69 mmol/l (+0.59, +0.79; p<0.0001) and current NFV use was associated with an additional total cholesterol *decrease* of -0.11 (-0.22, 0.00; p=0.05). Thus, compared to someone not receiving a PI, the effect of receiving NFV in the non-linear age model is a higher level as the child is currently receiving a PI, but this is offset with an average lower level as a result of receiving NFV rather than another PI. As NFV was by far the most common PI received amongst children in this cohort (see Table 8.1), perhaps these results are not so disparate as they first seem.

Table 8.5 – Factors associated with total cholesterol levels in HIV-positive children (assuming age is a linear effect)

		Unadjusted analysis			Adjusted analysis		
		Mean change (mmol/l)	**95% CI**	**p-value**	**Mean change (mmol/l)**	**95% CI**	**p-value**
Age	*Per year older*	+0.03	+0.01, +0.05	0.0005	0.00	-0.02,+0.02	0.76
Gender	*Male vs. female*	-0.03	-0.21, +0.15	0.75			
Age/gender interaction		+0.02	-0.02, +0.06	0.35			
Ethnicity	*African vs. other*	-0.16	-0.36, +0.04	0.11			
Ethnicity/age interaction		+0.03	-0.01, 0.07	0.20			
Height z-score	*Per 1 S.D. higher*	+0.02	-0.02, +0.06	0.32	-0.07	-0.12, -0.03	0.002
Height/age interaction		-0.01	-0.02, -0.01	0.001			
Weight z-score	*Per 1 S.D. higher*	+0.13	+0.09, +0.17	<0.0001	+0.11	+0.06, +0.15	<0.0001
Weight/age interaction		-0.01	-0.02, -0.01	0.007			
CD4 percentage	*Per 1% higher*	-0.02	+0.02, +0.02	<0.0001	+0.01	+0.01, +0.01	<0.0001
CD4 percentage/age interaction		+0.00	-0.00, +0.00	0.13			
Exposure to mono/dual therapy	*Yes vs no*	+0.19	-0.02, +0.40	0.07			
Monotherapy/age interaction		+0.01	-0.04, +0.05	0.71			
Current use of ART	*Yes*	+0.70	+0.64, +0.76				
Current use of PIs	*Yes*	+0.54	+0.47, +0.60	<0.0001	+0.69	+0.59, +0.79	<0.0001
Current use of NNRTIs	*Yes*	+0.35	+0.30, +0.41	<0.0001			
Current use of NFV	*Yes*	+0.34	+0.25, +0.43	<0.0001	-0.11	-0.22, 0.00	0.05
Current use of NVP	*Yes*	+0.37	+0.28, +0.45	<0.0001	+0.49	+0.41, +0.58	<0.0001
Current use of EFV	*Yes*	+0.31	+0.24, +0.39	<0.0001	+0.52	+0.44, +0.60	<0.0001
Ever exposed to ART	*Yes*	+0.75	+0.66, +0.84	<0.0001			
Effect of additional ART exposure	*>1 year*	+0.21	+0.14, +0.28	<0.0001			
	>2 years	+0.03	-0.05, +0.10	0.52			
	>3 years	-0.03	-0.11, +0.05	0.46			

		Unadjusted analysis			Adjusted analysis		
		Mean change (mmol/l)	**95% CI**	**p-value**	**Mean change (mmol/l)**	**95% CI**	**p-value**
	>4 years	+0.02	-0.05, +0.09	0.60			
Ever exposed to PIs	*Yes*	+0.63	+0.54, +0.72	<0.0001	+0.27	+0.16, +0.38	<0.0001
Effect of additional PI exposure	*>1 year*	+0.15	+0.06, +0.24	0.002	+0.10	+0.01, +0.19	0.03
	>2 years	+0.02	-0.08, +0.12	0.70	+0.03	-0.06, +0.12	0.62
	>3 years	-0.03	-0.13, +0.07	0.54	-0.05	-0.15, +0.04	0.26
	>4 years	-0.07	-0.17, +0.03	0.19	-0.02	-0.12, +0.08	0.67
Ever exposed to non-PI ART	*Yes*	+0.48	+0.40, +0.56	<0.0001	+0.04	-0.05, +0.14	0.36
Effect of additional non-PI exposure	*>1 year*	+0.14	+0.06, +0.22	0.0007	+0.07	0.00, +0.15	0.06
	>2 years	-0.06	-0.15, +0.03	0.20	-0.04	+0.12, +0.05	0.42
	>3 years	-0.13	-0.23, -0.03	0.01	-0.04	-0.14, +0.06	0.43
	>4 years	+0.08	-0.02, +0.19	0.12	0.00	-0.11, +0.10	0.93

Using mixed effects models, with age as the underlying time variable, which is included in each model

8.6 Discussion

In this chapter, I have considered the use of multivariable fractional polynomials to investigate non-linear relationships between variables. I found that total cholesterol levels in HIV-positive children were higher amongst those currently using PI- and NNRTI-containing ART, and with longer cumulative exposure to PI-containing ART. I also found higher total cholesterol levels were associated with higher CD4 cell percentage, higher age-adjusted weight and lower age-adjusted weight.

A recent study by the Swiss HIV Cohort Study used fractional polynomial models to investigate the impact of PI- and NNRTI-based HAART on lipid levels, BMI and blood pressure in HIV-positive adult men [438]. They included 556 men with pre-HAART measurements of the coronary risk factors, and at least one subsequent set of measurements in the 36-month period after starting HAART. They used multi-level models, including all pre-HAART measurements in the 6 months prior to treatment, and all measurements in the first three years of HAART. The authors found that total cholesterol levels increased rapidly in the first three months of HAART, before increasing more slowly from then on. This is consistent with the results for HIV-positive children found in this chapter. The authors also found that total cholesterol increases amongst those on PI-based HAART were slightly higher than those on NNRTI-based regimens, which is also consistent with my findings. A similar pattern was observed for HDL cholesterol levels, although here increases were greater amongst those on NNRTI-based regimens. Changes in triglycerides were larger amongst those receiving PI-based regimens.

A further study, from the European Collaborative Study (ECS) has used fractional polynomial models to investigate HIV-1-RNA viral load levels in HIV-positive children [439]. They found that HIV-1-RNA levels peaked in the first three months of life, gradually declining thereafter; they also found that girls had higher levels than boys, although this did not translate into a different rate of clinical progression. Multiple polynomial fractional models have also been used in several studies outside of the HIV field [440-460]. Examples include investigating the ability of the Revised Trauma Score for predicting in-hospital mortality [440], the association between C-reactive protein levels and age in Aboriginal Australians [441] and the calculation of age-specific reference intervals for carotid-femoral pulse wave velocity [445]. Thus, these methods have been widely used in a number of real-life situations.

Two studies have investigated both multiple fractional polynomial models and spline regression curves. Polesel et al [446] estimated the association between alcohol use with upper aero-digestive tract cancers, and Moore et al [443] studied the association of the Glasgow Coma Score with in-hospital mortality. In both studies, the authors concluded that spline regression curves resulted in a better model fit. However, as well as selecting the best-fitting model, it is important to consider the biological plausibility of any associations between variables suggested by the model. Results obtained from regression models in which spline curves are fitted, such as those shown in Figure 8.1, might suggest associations that may not be biologically appropriate. A simulation study by Roytson and Sauerbrei [461] concluded that multiple fractional polynomial models led to more realistic models being obtained than when using spline curves. Furthermore, the associations found in this chapter also appeared plausible suggesting that, in this example at least, a multiple fractional polynomial model was a reasonable approach to take. Furthermore, as they are also sensitive to the specific dataset on which they are calculated, it is also important to investigate whether similar results are found in more than one study.

An advantage of the use of fractional polynomial models is that they are easy to apply. A systematic approach is taken to choosing the most appropriate model, and a finite number of curves are considered. When nested models are considered, a statistical test can be performed to compare candidate models. Nonetheless, there is still great flexibility in the form of the association between variables, and so most associations between variables can be sufficiently described using this approach. A further advantage is that multiple fractional polynomials can be applied to any regression model. Therefore, they have a wide applicability.

An attractive property of linear associations between variables is that they are easy to interpret, as a change of one unit in the explanatory variable is associated with a change in the dependent variable, or in the risk of an event occurring. When associations between variables are non-linear, it is not necessarily so easy to intuitively interpret the association between variables, particularly when two terms are required. Consequently, it is useful to plot the relationship between the variables, such as via plots of fractional residuals. This can give an understanding of how the two variables interact. Additionally, one could calculate the estimated change in the dependent value (or increase in the risk of an event occurring) at specific values of the independent variable. Nonetheless, multiple fractional polynomial models do not quantify in a simple way the impact of a unit change in the explanatory variable and so must be explained carefully. Interpretation can be even more difficult when the dependent variable of

interest is binary, and non-linear associations with the log odds of an event occurring are obtained. One possible solution could be to consider the independent variable as a categorical variable. In this situation, interpretation of the results obtained is easier. The disadvantage, however, is that one must choose the appropriate cut-off for each group. This can lead to a number of potential biases, and furthermore subtle associations can be missed. In my model in this chapter, I decided to consider ART exposure as a categorical variable. Although this may have led to potential biases as described above, the categories chosen were defined *a priori*, and appeared suitable given the results of the descriptive analyses in Figures 8.7 to 8.9.

The results of this chapter found that there was evidence of an association between age and total cholesterol levels. This association was only apparent when age was fitted as a non-linear term, and so would have been missed otherwise. This highlights the importance of investigating whether non-linear associations are appropriate, particularly in settings such as this where there is a biologically plausible reason for non-linearity. The model suggests that total cholesterol levels increase until puberty, and decline gradually thereafter. This largely corroborates the association observed in other studies [421;422]. However, other studies have generally found a decrease immediately after infancy, before levels increased again as puberty began., which has not been captured when using multiple fractional polynomials here. Therefore, perhaps a more complicated model is required to fully capture this association, although this must be balanced against the biological plausibility of any model obtained. Nonetheless, the primary aim of the study was to investigate when there was a relationship between ART use (and PI use in particular), and total cholesterol. It is interesting to note that relationship between ART use and total cholesterol was similar regardless of whether age was controlled for as a linear or non-linear association. Therefore, in this example, if we were only interested in controlling for the effect of age, consistent results for other variables would be obtained regardless of whether a final model was chosen that included a linear or a non-linear age term.

There is always a risk when fitting non-linear associations between variables of over-fitting the model, and obtaining an association that, although fits well to the current dataset, will not do so in subsequent studies. Therefore, one must think carefully before fitting non-linear associations. In this chapter, I was considering an association – that between age and total cholesterol levels in children, that had previously been shown to be non-linear [421;422]. Therefore, in this situation, I felt it was reasonable to consider a non-linear relationship between variables.

The results of this study imply that receiving ART is associated with increases in total cholesterol levels, confirming that of previous studies in both children [424;428-430] and adults [217;268;273]. Current use of a PI, NVP or EFV was associated with increased levels. Those receiving NFV had, on average, lower total cholesterol than those receiving other PIs by 0.11 mmol/l, although levels were still higher than in those not receiving any PIs. The current use of NVP and EFV was associated with a comparable increase in total cholesterol levels compared to current NFV use, and a slightly smaller increase compared with PIs other than NFV. Other studies have found that greater total cholesterol increases are associated with PI-containing regimens compared to non-PI containing regimens[207;217;252].

Independently of current ART use, cumulative exposure to PI-containing ART was also associated with increased total cholesterol levels. This increase in levels appeared to occur in the first year of treatment, after which levels remained consistent. This corroborates the findings from adult studies [438;462]. Thus, individuals who have received PI-containing ART remain exposed to raised total cholesterol levels after a short time of exposure, and thus are likely to have a long-term increased cardiovascular risk [418;419]. In contrast to current use, cumulative exposure to non-PI containing regimens was not associated with total cholesterol levels, in agreement with other paediatric studies [424;429].

The study by Rhoads et al [428] demonstrated that, as total cholesterol levels increased in HIV positive children in the first years of ART, so too did HDL cholesterol levels. Furthermore, they appeared to be below the normal range prior to treatment. Thus, as HDL cholesterol is a component of the total cholesterol level, an increase in total cholesterol may to some extent reflect a beneficial effect of ART. ART-associated HDL cholesterol increases have also been found in adult studies, and has been particularly associated with NNRTI use [438]. Thus, the increase in total cholesterol may to some extent reflect a positive outcome, as a result of increased HDL cholesterol levels. Unfortunately, it was not possible to obtain cholesterol fractions from both hospitals, and I was therefore unable to investigate this further in the present study. It would be interesting in future studies to investigate this in more detail.

Other factors found to be associated with higher total cholesterol levels were a higher CD4 percentage, a lower age-adjusted height z-score and a higher age-adjusted weight z-score. Higher CD4 percentages reflect an improved immune status and general well being, which may include total cholesterol levels. As higher CD4 percentages are also related to successful antiretroviral therapy, this finding may also

in some way reflect the association between ART use and total cholesterol levels. However, there is evidence that, in adults, the total cholesterol level decreases with the CD4 count. A link between increased weight and higher long-term cardiovascular risk in children has been shown [417;418], and therefore an association between weight z-score and total cholesterol levels is perhaps to be expected. The association between lower height z-scores and higher total cholesterol levels was only observed after adjusting for weight z-score. Thus, a lower height z-score for two individuals of the same weight means that a lower height z-score is associated with a higher BMI, and thus an association with higher total cholesterol levels is to be expected.

The potential implications of exposure to raised lipid levels from childhood for potentially a great number of years if ART is required may potentially lead to a long-term increase in the risk of cardiovascular disease. Other interventions such as lifestyle changes and the use of statins may in the long-term lead to reductions in the risk of cardiovascular disease in HIV-positive children.

8.7 Summary

To summarise this chapter, multivariable fractional polynomials are a method for investigating non-linear relationships between variables. They are simple to use and thus can be useful in an applied setting. Although the form of the non-linear relationship obtained from a fractional polynomial model may not lead to an intuitive understanding of the relationship between variables, plots of fractional residuals can illustrate the association. In our example of total cholesterol levels in HIV positive children, I found that, although there was evidence of a non-linear association between age and total cholesterol levels and the association between the two variables was only apparent when age was fitted as a non-linear term, the relationship between ART use and total cholesterol was similar regardless of whether age was controlled for as a linear or non-linear association.

Chapter 9 – Concluding remarks

9.1 Summary of main findings

As a result of the dramatic declines in the prevalence of HIV-related mortality and morbidity with the introduction of HAART in around 1996 [88;463], there has been a shift from considering HIV as a terminal disease to that of a chronic condition. Whilst not forgetting the benefits of HAART, however it is important also to be aware of any potential toxicities and side effects that may be associated with this treatment. In order to make informed choices with regard to these drugs, it is important to accurately quantify the magnitude of any HAART-related toxicities, and also to compare the associated risks of different antiretrovirals in an unbiased manner. As a result, the aim of my thesis was to investigate the methodological issues and potential biases present when investigating HAART-related toxicities, and to consider methods for assessing the prevalence and incidence of these toxicities which are least biased, and which are able to be used to accurately compare antiretrovirals and different demographic populations.

There are a number of different toxicities that may be associated with ART, and a number of different antiretrovirals each associated with different toxicities. Thus, consideration of HAART regimens as a whole group, regardless of the specific antiretrovirals included in the regimen, may not always provide the most useful information. Nonetheless, in my thesis I have attempted to investigate methods for estimating HAART-related toxicity that apply to all laboratory toxicity markers, as well as considering more specific issues relating to individual markers. I have also investigated the occurrence of treatment-limiting toxicities, and hope that the methods employed in Chapter 6 that consider the best way to account for non-toxicity related treatment discontinuations, could be extended if one was interested in a particular toxicity. For example, if one wished to investigate the occurrence of peripheral neuropathy that led to the discontinuation or switch of an antiretroviral, we could use the methods discussed in Chapter 6, accounting for patients switching for non-peripheral neuropathy reasons in the same way that I accounted for non-toxicity reasons in this Chapter. However, the issue of informative censoring and the interpretation of results that account for that when patients discontinue antiretrovirals for reasons other than the one of interest remains.

I began in Chapter 4 by considering the potential biases associated with changes over time in the frequency of monitoring of laboratory markers used to measure HAART-related toxicities. This chapter demonstrated that there has been a marked increase in more recent calendar years in the frequency of monitoring of laboratory markers. As there have also been changes in the demographic population starting HAART at the Royal Free Hospital over the same time period, and as new antiretrovirals have been introduced, investigation of the impact of these issues is likely to be biased by differential monitoring. Hopefully, as time passes, this will become less of an issue, as patients are monitored more routinely. However, the use of endpoints that are least affected by differential rates of monitoring will remain optimal. My simulation model suggested that there is great variability in the extent to which different endpoints may be affected by biases associated with frequency of monitoring. Therefore, choice of an endpoint, such as the first measurement in a reasonably large window, (e.g. six months to one year after starting HAART), and exclusion of those without a measurement in this window appears to be a reasonable endpoint. However, use of the endpoints recommended here with a missing=excluded endpoint must be weighed up against any biases this may introduce if those excluded are those who were most likely to have experienced toxicity.

Chapter 5 considered the most appropriate cut-off to use to define the occurrence of a toxicity. I focused here on two particular toxicities – hypercholesterolaemia and hepatotoxicity. I found that the choice of endpoint led to differences in the estimated prevalence of toxicity. Thus, studies using different cut-offs are likely to not be easy to compare. Identifying the most clinically relevant cut-offs is not straight forward. There was limited switching as a result of lipid abnormalities, as this was so infrequently given as a reason for stopping an antiretroviral (just three occasions in the first year of HAART), particularly in comparison to the number of patients who experience high total cholesterol levels. Therefore, identification of what is thought of as a clinically relevant high total cholesterol level in this way is not possible. Guidelines, such as those issued by the UK Department of Health, create target levels for total cholesterol levels to be below, and studies have typically used a cut-off of 6.2 or 5.5 mmol/l, although there is variation in this. However, it is important to also ensure that high total cholesterol levels do not merely reflect high pre-HAART level, and therefore incorporating information on the change in total cholesterol levels may be appropriate. This led to choosing a cut-off of >5.5 mmol/l with at least a 1 mmol/l increase from baseline as a candidate endpoint. When considering hepatotoxicity events, it was immediately clear that the prevalence of these events, in the first year of HAART at least, was low. Furthermore, when considering those who discontinued or switched antiretrovirals as a

result of abnormal LFTs, this indicated that clinicians were intervening before many of the criteria for appropriate cut-offs were met. Furthermore, many of the patients who were discontinuing an antiretroviral for hepatotoxicity reasons had high levels at baseline. Therefore, choice of an endpoint with a lower threshold than those commonly used in the published literature, such as 2.5 times the upper limit of normal, or preferably changes from pre-HAART levels may be the most appropriate choice when considering hepatotoxicity. As for both hypercholesterolaemia and hepatotoxicity endpoints there was evidence supporting the choice of *change* in total cholesterol or AST/ALT levels as an appropriate endpoint, it may potentially be reasonable to consider this type of endpoint for other laboratory markers.

I applied the suggestions of Chapters 4 and 5 with regard to the most appropriate endpoint to use in analyses in Chapter 6 when considering the impact of the CD4 cell nadir on the occurrence of toxicities. I found that the endpoints suggested by Chapters 4 and 5 were easy to calculate and apply. However, use of these endpoints in this analysis meant that the number of events was very low. As I was considering here all patients starting HAART from 1st January 1997 until 31st December 2004, the number experiencing an event in certain subgroups, (e.g. those starting PI-containing HAART or considering specific demographic groups), is likely to be even lower. Thus, choice of an endpoint that is more affected by frequency of monitoring, but is the 'next best' candidate, may be a more feasible approach. Here, the endpoint chosen would be any high measurement in the first year of HAART, taking a missing=failure approach if the value is missing. Alternatively, a longer period of follow-up could be chosen, or a smaller change in the laboratory marker, although this may lead to identifying patients as having experienced an event when in fact they have only experienced random variation in the laboratory marker.

Chapter 7 considered an issue that is particularly relevant when considering observational studies, and that will apply to a number of settings; that of unmeasured confounding. Although randomised controlled trials will and should remain as the gold-standard way of assessing causal effects, there remain situations, such as those identified in Chapter 2, where this is not possible. In this case the best available evidence is likely to come from observational studies, and any methods that may be able to go some way to accounting for unmeasured confounding are likely to be useful. Whilst the theory behind Sample Selection models is not necessarily intuitive and the Inverse Mills Ratio is not easy to interpret, these models are easy to apply in real life situations. Furthermore, the results of my simulation study imply that they are able to account for unmeasured confounding, at least in the simulated situations I investigated.

However, when applying these models to a real-life situation, I found that there was great imprecision in the estimates obtained. The published literature that have used these methods have generally considered cohorts with larger patient numbers. If a limitation of this method is that a large sample size is required in order to obtain precise estimates then this may limit the usefulness of the method in certain situations. Furthermore, when applying the models, one must make a number of assumptions with regard to the known and measured factors associated with treatment allocation only, the response variable only and those that are confounders. Although in my simulation model this information was available, in real life situations it may be harder to ascertain which of these situations hold for each measured variable.

Finally, Chapter 8 considered the use of multiple fractional polynomials to investigate whether non-linear relationships between variables exist. In this chapter I have concentrated on any ART use, including exposure to any mono or dual therapy prior to the start of HAART. I have also included follow-up on patients from the time of their first total cholesterol measurement, regardless of their antiretroviral treatment history at that time. Furthermore, this analysis also included patients under follow-up during which time they did not receive any antiretroviral therapy, so that they could act as their own controls, as well as controls to those who were receiving ART at the time. In this way, the type of analysis performed in Chapter 8 is quite different to that of previous chapters. Although considering all patients from the time of first starting HAART means that any analyses performed are on a 'cleaner' dataset, and it may be easier to disentangle particular associations, use of an approach such as the one taken in Chapter 8 enables a greater length of follow-up to be included and to include a naïve or not currently receiving ART control group in the analyses.

The issue of non-linear associations between variables is an important one, not only so that we appreciate more fully the associations between variables, but also to avoid incorrect conclusion about the associations between variables. However, there are an infinite number of potential relationships between variables, and therefore to limit this pool of potential models and to provide a systematic method for choosing the most appropriate association is of benefit. I found that applying multiple fractional polynomials to the Great Ormond Street Hospital/St Mary's Hospital dataset was easy, particularly as Royston and Altman have developed a procedure for the statistical package Stata, *mfp*. In this example, although there was a non-linear association between age and the total cholesterol level in children, this association did not impact on the observed association between total cholesterol and other explanatory factors in the model, particularly ART treatment use. It is useful to produce plots of the estimated

associations between variables when using multiple fractional polynomial models, as the formulae for these associations are sometimes complex and therefore not easy to interpret. The *mfp* procedure presents these plots. Multiple fractional polynomials can be used in a number of situations to investigate whether linear associations between dependent variables and potential explanatory variables exist. For example, they could be used to investigate whether there was an association between the length of exposure to HAART and haemoglobin levels. Furthermore, they are flexible in that they are able to characterise a number of different relationships, such as logarithmic relationships and polynomial relationships. Thus, they are a useful tool when investigating the association of factors with the occurrence of antiretroviral-related toxicity.

9.2 Plans for future work

In this thesis, I have concentrated on HIV positive patients who are antiretroviral naïve and starting HAART for the first time. Furthermore, I have focused on the short-term toxicities observed in the first year of HAART. Although many of the issues investigated in this thesis here with regard to potential bias present can extend to the longer-term situation, and to treatment experienced patients, it would be interesting to investigate these issues further. I have also, with the exception of Chapter 9, investigated whether potential biases are present in only a single dataset, the Royal Free Hospital database. Although many potential biases will be present in all HIV observational datasets, of course each database will be slightly different. Therefore, it would be interesting to repeat some of the analyses performed here in different datasets to investigate whether the same biases are present.

I have considered the issues here in the specific setting in which we are considering the occurrence of antiretroviral-related toxicities. However, many of the underlying ideas may also be applicable to other areas of research in the HIV field, and to research of other diseases. It would be of interest to implement the methodology presented here in other situations.

I have used a treatment discontinuation endpoint for toxicity-related events. The limitations of this approach have been discussed in detail in Chapter 4. Otherwise, I have used surrogate laboratory markers, such as total cholesterol and AST levels, to investigate HAART-related toxicity. Unfortunately, the Royal Free Cohort does not collect information on clinical toxicity events such as lipodystrophy and peripheral

neuropathy unless they are treatment limiting, and there is limited power, even in a large teaching hospital cohort such as the Royal Free, to study life-threatening clinical events such as pancreatitis and myocardial infarction. Therefore, I have not been able to investigate the issues associated with estimating the prevalence of these events. Although many of the potential issues were raised in Chapter 2, it would be interesting to investigate these in more detail, perhaps in as part of a cohort collaboration.

Finally, although this thesis has mainly concentrated on the methodological issues and potential biases that occur when investigating antiretroviral-related toxicities, I have tried to consider endpoints and methodologies that are simple to apply and thus could be used easily in general clinical research. Although it is important to always consider the clinical relevance of studies, and to ensure that they are easily interpretable, it is clearly also important to ensure that the findings of any study are methodologically sound and all attempts are made to minimise potential biases, in order to obtain results that we believe to be correct. Therefore, it would also be of interest to apply the methods here to answer clinically relevant questions that will provide evidence for the best possible clinical care for patients.

9.3 Summary and concluding remarks

When carrying out research in observational datasets we are never completely certain that all potential biases have been completely removed. Nonetheless, there are often situations in which the only evidence on a particular issue will be obtained from observational datasets and thus it is important to do everything we can to minimise these biases and therefore obtain results in which we have confidence. I have investigated a method to account for unmeasured confounding, the best method to account for differential frequency of monitoring, the most clinically relevant surrogate endpoints when considering laboratory markers, and the application of non-linear regression models for situations when non-linear relationships between variables may exist. I hope that this will contribute to research investigating the best ways to measure the occurrence of antiretroviral related toxicities.

Appendix A – Data collection forms for Royal Free Cohort

A.1 First visit form

NEW PATIENT REGISTRATION

Hospital No. _____ Date of Visit : ___/___/___

Surname : _____ First Name : _____

Date of birth : ___/___/___ Sex : Male ☐ Female ☐

Is this patient a transfer? Yes ☐ No ☐ If Yes where from? _____

Country of Birth : _____

Stage 1 interviewer : _____ Is a stage 2 interview required : Yes ☐ No ☐

Patients Address : _____

_____

_____

Full Postcode : _____

Contact Numbers : Home _____ Mobile _____

Work _____ E-mail _____

Ethnic Group :

White ☐ Black African ☐ Black Caribbean ☐ Black Other (Please Specify) _____

Indian ☐ Pakistani ☐ Bangladeshi ☐ Other (Please Specify) _____

GP DETAILS :

Name : _____

Address : _____

_____

Telephone : _____

Is GP aware of Diagnosis : Yes ☐ No ☐

Can he/she be contacted ? Yes ☐ No ☐

NEXT OF KIN DETAILS :

Name : _____

_____

_____

Telephone: _____

Aware of Diagnosis : Yes ☐ No ☐

Patient signature : _____ Date : _____

Last centre of HIV/AIDS Care? (if applicable) :

UCL / Mortimer Mkt ☐	Kobler / C & W ☐	St Mary's ☐	St Thomas's ☐
Bart's / Royal London ☐	Whittington ☐	King's ☐	Guy's ☐
North Middlesex ☐	Barnet General ☐	Chase Farm ☐	Hammersmith ☐
Blood Transfusion Service ☐	Other (specify) _____		

Can last centre be contacted for past medical history? Yes ☐ No ☐ Hospital No. _____

Patient Signature : _____ Date : _____

HHA917

CIRCUMSTANCES LEADING TO PRESENTATION

Date of first NEGATIVE antibody test : _____

Where was this test performed?

Royal Free SDTC ☐	Royal Free Marlborough ☐	Royal Free Antenatal ☐	
UCL / Mortimer Mkt ☐	Kobler / C & W ☐	St Mary's ☐	St Thomas's ☐
Bart's / Royal London ☐	Whittington ☐	King's ☐	Guy's ☐
North Middlesex ☐	Barnet General ☐	Chase Farm ☐	Hammersmith ☐
Blood Transfusion Service ☐	Other (specify) _____		

Date of last NEGATIVE antibody test : _____

Where was this test performed?

Royal Free SDTC ☐	Royal Free Marlborough ☐	Royal Free Antenatal ☐	
UCL / Mortimer Mkt ☐	Kobler / C & W ☐	St Mary's ☐	St Thomas's ☐
Bart's / Royal London ☐	Whittington ☐	King's ☐	Guy's ☐
North Middlesex ☐	Barnet General ☐	Chase Farm ☐	Hammersmith ☐
Blood Transfusion Service ☐	Other (specify) _____		

Date of first POSITIVE antibody test : _____

Where was this test performed?

Royal Free SDTC ☐	Royal Free Marlborough ☐	Royal Free Antenatal ☐	
UCL / Mortimer Mkt ☐	Kobler / C & W ☐	St Mary's ☐	St Thomas's ☐
Bart's / Royal London ☐	Whittington ☐	King's ☐	Guy's ☐
North Middlesex ☐	Barnet General ☐	Chase Farm ☐	Hammersmith ☐
Blood Transfusion Service ☐	Other (specify) _____		

Reason for test :

Symptoms ☐	Known positive partner ☐	Risky Behaviour ☐	Antenatal ☐
Blood Donor ☐	Insurance/Visa Screen ☐	Confirmation of known positive ☐	

Other (specify) : _____

Probable Route of Infection

Patient presumed infected in the UK? Yes ☐ No ☐ Not Known ☐

If No or Not known in which countries? _____

Sexual relations between men Yes ☐ No ☐

Does this patient believe themselves to be infected through oral sex only ? Yes ☐ No ☐

Sexual relations between men and women

Has this patient had sex with - Bisexual male Yes ☐ No ☐ Not Known ☐

- Injecting Drug user Yes ☐ No ☐ Not Known ☐

Partner presumed heterosexually infected Yes ☐ No ☐ Not Known ☐

If Yes partners likely country(ies) of infection _____

Injecting Drug Use (sharing)

Year first injected _____ Year last injected _____

Does the patient have protected sex (condoms) : Always ☐ Sometimes ☐ Never ☐

Patient defined Sexual Orientation?

Homosexual ☐ Heterosexual ☐ Bisexual ☐ Other : _____

Marital Status

Single ☐ Married ☐ Reg. Cohabitator Male ☐ Reg. Cohabitator Female ☐

Widower ☐ Partner died ☐ Separated / Divorced ☐

Does patient have a partner currently ? Yes ☐ No ☐

If YES is partner Male ☐ Female ☐

Is partner HIV positive ? Yes ☐ No ☐ First name _____

If Yes does partner have AIDS ? Yes ☐ No ☐ Unknown ☐

Does patient have any children ? Yes ☐ No ☐

If YES how many children ? 1 ☐ 2 ☐ 3 ☐ 4 or more ☐

Are any HIV positive ? Yes ☐ No ☐

If YES where are they being treated ? _____

Comments e.g. Housing, Employment, Methods of contraception etc

---

Current History

Cigarette smoking & alcohol consumption

Does patient smoke? Yes ☐ No ☐ per day ? _____ Number of years _____

Does patient drink ? Yes ☐ No ☐ units per week _____ Number of years _____

Does patient use recreational drugs ? Yes ☐ No ☐ Which ? _____

Past Medical History (operations):

Has the patient ever been vaccinated for :

BCG ☐ Which year _____

Hep A ☐ Which Year _____

Hep B ☐ Which Year _____

Known Allergies : _____

Pre-therapy Viral Load (if known) Value _____ Date _____

Lowest ever CD4 count (if known) Value _____ Date _____

Family History : Specifically Diabetes, Heart Disease, MI, Stroke, DVT, Cancer

Diagnosed before age 50?

Mother: _____ ☐

Father: _____ ☐

Brother(s): _____ ☐

Sister(s): _____ ☐

Other (specify): _____ ☐

All Current Medication

Drug	Dose	Purpose

Known allergies :

Antiretroviral adherence : _____ ?

Past Antiretrovirals

Drug	Dose	Started	Stopped	Reason

Complete for any NEW or PREVIOUS AIDS Diagnosis

Please tick the AIDS indicator disease(s) and give month and year of diagnosis

AIDS Indicator Disease	Diagnosis Definitive	Diagnosis Presumptive	Date Month/Year
Candidiasis: trachea, bronchi, or lungs	☐	NA	/
Candidiasis: oespohageal	☐	☐	/
Cervical carcinoma, invasive	☐	NA	/
Coccidioidomycosis : expulmonary	☐	NA	/
Cryptococcosis : expulmonary	☐	NA	/
Crytosporidiosis with diarrhoea > 1 month	☐	NA	/
Cytomegalovirus retinitis	☐	☐	/
CMV disease not liver, spleen or nodes	☐	NA	/
Encephalopathy (dimentia) due to HIV	☐	NA	/
Herpes simplex : ulcers > 1 month or bronchitis, pneumonitis, oesphagitis	☐	NA	/
Histoplasmosis ; disseminated/extrapulmonary	☐	NA	/
Isosporiasis with diarrhoea for > 1 month	☐	NA	/
Kaposi's sarcoma	☐	☐	/
Lymphoma, Burkitt's or equivalent term	☐	NA	/
Lymphoma, immunoblastic or equivalent term	☐	NA	/
Lymphoma, primary in brain	☐	☐	/
Mycobacterium avium: extrapulmonary	☐	☐	/
M.tuberculosis : pulmanary	☐	☐	/
M.tuberculosis : extrapulmonary	☐	☐	/
Mycobacterium of other or unidentified species disseminated	☐	☐	/
Pneumocystis carinii pneumonia	☐	☐	/
Pneumonia recurrent within 12 months	☐	☐	/
Progressive multifocal leukoencephalopathy	☐	NA	/
Salmonella septicaemia, recurrent	☐	☐	/
Toxoplasmosis of the brain	☐	☐	/
Wasting syndrome due to HIV	☐	☐	/

Clinical Examination	Normal	Abnormal findings
General	☐	
Cardiovascular	☐	
Skin	☐	
Chest	☐	
Mouth	☐	
Abdomen	☐	
Lymph nodes	☐	
Neurological /Fundi	☐	

CLINICAL STAGE

Asymptomatic ☐ Symptomatic ☐ AIDS ☐

SUMMARY OF CASE

1. _____

2. _____

3. _____

4. _____

5. _____

6. _____

Letter to GP? Yes ☐ No ☐ Next Visit _____ Weeks

Doctor's Name : _____ Referrals : _____

Post : _____

Hepatitis Clinic ☐

Womens Clinic ☐

Contact blp/ext. : _____

Lipid Clinic ☐

A.2 Patient clinic visit form

Name	Visit date:	☐ Booked
Hospital No.	Weight _____ kg	☐ Unbooked

History and Examination

	Date
VL:	
CD4:	
CD4 %:	

☐ Current trial

Reasons for stopping:

1 Failure VL ↑
2 Failure CD4 ↓
3 Study change
4 Rash
5 Nausea/vomiting
6 Diarrhoea
7 Mouth ulcers
8 Abdominal pain
9 ↑ LFTs
10 ↑ Amylase
11 Pancreatitis
12 ↑ Glucose/diabetes
13 ↑ Lipids
14 Fat wasting (LD)
15 Fat accumulation (LD)
16 Lactic Acidosis
17 Myositis
18 Renal problem
19 Anaemia
20 Malaise/fatigue
21 CNS effects
22 Headache
23 Peripheral Neuropathy
24 Allergic reaction
25 Drug interaction
26 Poor compliance
27 Rationalisation (eg Trizivir)
28 Patient choice (no a/es)
29 Failure ↑ resistance
30 Other (specify)

New diagnoses	Date	Current drug trial? Y ☐
AIDS :		Short title :
Other HIV :		RFH no :
Other (eg CVD):		Subject no :
		Visit / week :

Obs ☐
Bloods ☐
QoL ☐

Antiretroviral Medication ☐ *No change* Proportion of pills taken in the last month ☐ %

Drug with dosage	Started	Stopped	Prescribed	Reasons (codes above)		Date stopped
	○	○	○			
	○	○	○			
	○	○	○			
	○	○	○			
	○	○	○			
	○	○	○			

Other drugs with dosage ☐ *No change* Investigations:

Referral **GP letter ?** Y ☐ N ☐

Doctors name: **Next visit in** weeks

A.3 Data collection form

Demographic Data

Patient Name | Hospital number | DOB | Gender

Risk | Ethnicity

Visit Dates

First seen | Last seen | Tifer from

Date 1st +ve | Date last -ve | Death | Currently on trial

Specify

Smoking in past year: Unknown / Smoker / Non-Smoker at date:

Clinic Forms (since 1/1/05) | Prescriptions (since 1/1/05)

Treatment History

Drug	Date started	Date stopped	Reasons for stopping			Total daily dose

Clinical Diagnoses

Date	Diagnosis

Adherence

Date	good	average	poor	not st
	☐	☐	☐	☐
	☐	☐	☐	☐
	☐	☐	☐	☐
	☐	☐	☐	☐
	☐	☐	☐	☐

Admissions

Admission	Discharge	Reasons		

On PCP prophylaxis at date last seen?

	Yes	No
Co-trimoxazole	☐	☐
Dapsone	☐	☐
Pentamidine	☐	☐
Other	☐	☐

Date this form completed

Appendix B – Statistical Methods

B.1 Probability distribution functions of continuous random variables

B.1.1 The Normal Distribution

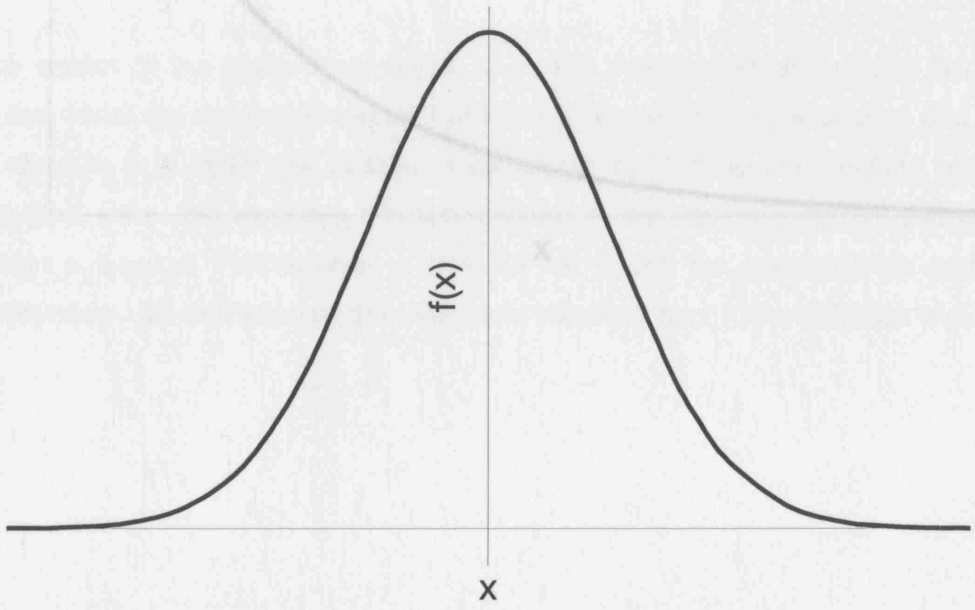
A random variable, X, with mean μ and standard deviation σ is Normally distributed (i.e. $X \sim N(\mu, \sigma^2)$) when the probability density function (pdf) of the variable, denoted by $\phi(x)$ or *f(x)*, is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty$$

and the cumulative density function (cdf), denoted by $\Phi(x)$, *F(x)* or *P(X < x)*, is given by:

$$P(X < x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \qquad -\infty < x < \infty$$

Figure B.1 – The probability density function of the Normal distribution with zero mean and standard deviation of one



B.1.2 The Exponential Distribution

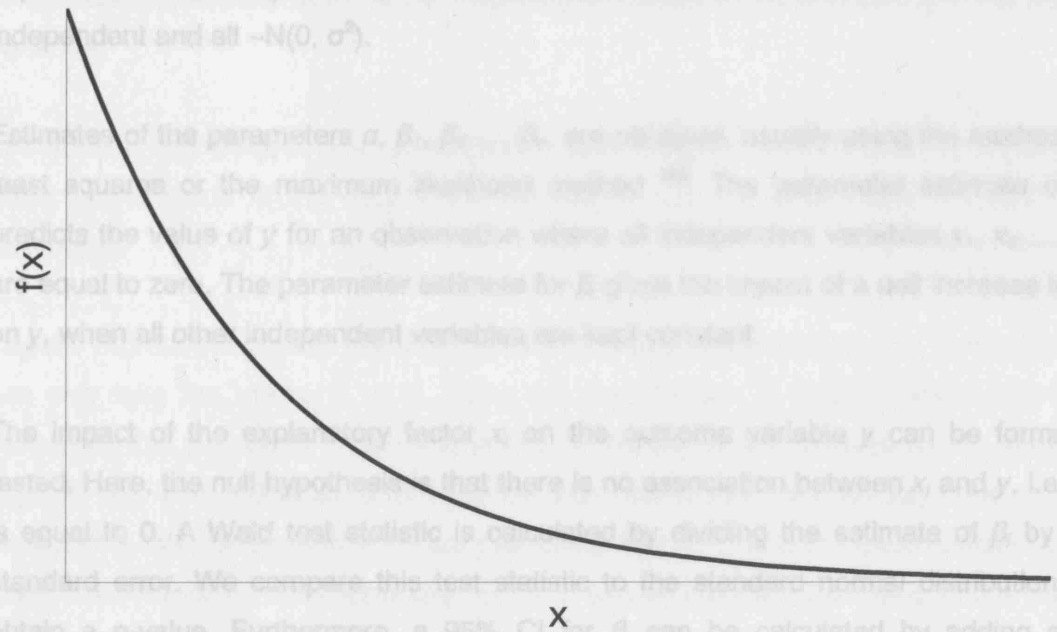
A random variable, X, with mean 1/λ and standard deviation 1/λ is Exponentially distributed (i.e. X ~ Exp(λ)) when the pdf of the variable, denoted by *f(x)*, is given by:

$$f(x) = \lambda e^{-\lambda x} \qquad x \geq 0$$

and the cdf, denoted by *F(x)* and *P(X < x)*, is given by:

$$P(X < x) = 1 - e^{-\lambda x} \qquad x \geq 0$$

Figure B.2 – The probability density function of the Exponential distribution with a mean of one and a standard deviation of one



B.2 Regression models

B.2.1 Linear regression

Linear regression models are employed when we wish to predict the value of a continuous variable, *y*, known as the *dependent* or *outcome* variable, from a set of *explanatory* or *independent* variables; x_1, x_2, x_2, ..., x_n. We estimate the magnitude of the effect of the independent variables on the dependent variable using a sample of observations from the population of interest. We assume that the association is of the following form:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_n x_n + \varepsilon$$

The random error term, ε, for each individual observation is assumed to be independent of all other error terms and each is from a normal distribution with expected value (mean) of zero, and with the same variance for each; i.e. ε are mutually independent and all ~N(0, σ^2).

Estimates of the parameters α, β_1, β_2,..., β_n, are obtained, usually using the method of least squares or the maximum likelihood method [389]. The parameter estimate of α predicts the value of *y* for an observation where all independent variables x_1, x_2,..., x_n are equal to zero. The parameter estimate for β_i gives the impact of a unit increase in x_i on *y*, when all other independent variables are kept constant.

The impact of the explanatory factor x_i on the outcome variable *y* can be formally tested. Here, the null hypothesis is that there is no association between x_i and *y*, i.e. β_i is equal to 0. A Wald test statistic is calculated by dividing the estimate of β_i by its standard error. We compare this test statistic to the standard normal distribution to obtain a p-value. Furthermore, a 95% CI for β_i can be calculated by adding and subtracting 1.96 multiplied by the estimate's standard error to the estimate of β_i.

B.2.2 Generalised Linear Models - Logistic regression

Generalised linear models are a general family of regression models that include linear regression and logistic regression. As in the linear regression example, we wish to investigate the impact of a set of independent variables on a dependent variable *y*. However, *y* no longer has to be a continuously distributed variable. Generalised linear models take the following form:

$$\mu = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + ... + \beta_n x_n + \varepsilon$$

where

μ = linear predictor

$\mu = f(\theta)$

$f(\theta)$ = link function

The random error term, ε, for each individual observation is assumed to be independent of all other error terms, each has an expected value of zero and the same variance.

In the linear regression model described in Appendix B.2.1, the link function is simply equal to the value of the dependent variable, y, and the error terms are normally distributed.

Logistic regression models are used when the dependent variable, *y*, is binary (i.e. it can only take the values 0 or 1). Here, the link function is logit(p), where p is the probability that *y* is equal to one and logit(p) is the log odds of p given by the formula $\log_e(p/[1-p])$. The estimates of β_i obtained from the model give the log odds of the impact of a unit increase in x_i on the probability that y=1, and the exponential of this gives the odds ratio. Wald tests and 95% CIs can be calculated similarly to the linear regression case.

B.2.3 Cox proportional hazards regression models

We are often interested in calculating the time to the occurrence of an event. When not all individuals experience the event (and perhaps have different lengths of follow-up) then we have censored data, as individuals are still at risk of experiencing the event in the future. We use survival methods to account for these censored data. For more details see Collett [464].

For each individual, we can calculate a hazard, *h(t)*, which is the instantaneous rate of having an event at any point in time, *t*, given that the individual has not experienced the event of interest up until this time point. Although we may not be interested in the value of the hazard itself, we are often interested in the hazard ratio – the multiplicative effect that a unit increase in a factor of interest has on the hazard. Cox proportional hazard regression models take the following form:

$$h(t) = h_0(t)\exp(\beta_1 x_1 + \beta_1 x_1 + ...\beta_n x_n) \qquad t > 0$$

Here, $h_0(t)$ is the baseline or underlying hazard. Cox proportional hazard regression models make no assumptions on the form of this hazard (and thus the models are semi-parametric). However, the models make the *proportional hazards* assumption; the multiplicative impact of factor x_1 on the hazard remains constant, regardless of the current time point.

The parameters β_1, β_2,... β_n, are estimated by maximising the partial likelihood. The estimate of $exp(\beta_i)$ gives the hazard ratio for factor x_i. We can construct hypothesis tests to assess the impact of factor x_i on the hazard and 95% confidence intervals similarly to that described for linear regression models.

B.2.4 Mixed effects regression models

All of the regression models described so far have considered the situation in which each observation is independent of others. However, in the situation in which samples are not independent, for example when more than one measurement on each individual has been taken, these regression models are no longer appropriate. Here, we must use mixed effects, or multilevel, models.

We assume here that the outcome of interest is continuous. In this model, each observation now has two (or more) levels – for example, the individual on which the measurement was taken (level 1) and the time at which the measurement was taken (level 2). We also have a first set of explanatory factors, x_1 to x_n, that are assumed to be associated with the same effect on the dependent factor, y, regardless of the time and individual under consideration, and a second set of explanatory factors, z_1 to z_p, which remain constant across all time points, but can randomly vary from an overall mean for each individual. Then the model takes the form:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \gamma_1 z_1 + \gamma_2 z_2 + ... + \gamma_p z_p + \varepsilon$$

we also assume:

ε is the overall error term for each individual at each time points, and are independent of each other and are identically Normally distributed with mean zero and constant variance

$\beta_1, \beta_2, ... \beta_n$, are fixed effects, and so the impact of x_i on y, estimated by β_i, is the same for all individuals at all time points.

$\gamma_1, \gamma_2, ... \gamma_p$, are random effects, and so, although there is an overall mean effect of the factor z_j on the dependent factor, y, for each individual the impact of the factor randomly varies from this mean with constant variance. However, for each individual, the effect of the factor z_j remains constant across all time points.

Again, Wald tests 95% CIs for the parameter estimates can be calculated as described above.

B.3 Sample Selection Models

B.3.1 Notation for formal derivation of Sample Selection model

The notation and derivation of formulae used in this section follows that of Crown et al [396]. I shall first introduce more formal notation for the situation described in Subsection 7.2.1

$i = 1,2,...,n$ subjects included in study

$\underline{X}_i$ = vector of known covariates for patient i

$\underline{X}^*_i$ = vector of known covariates for patient i that influence treatment allocation (subset of $\underline{X}_i$)

$$T_i = \begin{cases} 1 & \text{patient receives treatment A} \\ 0 & \text{patient receive treatment B} \end{cases}$$

$$Y_i = \begin{cases} 1 & \text{patient has a positive outcome} \\ 0 & \text{patient has a negative outcome} \end{cases}$$

B.3.2 Standard generalised linear model

Let us begin by considering the association in which we are interested in investigating, that is whether receiving regimen A rather than regimen B is associated with a different outcome, such as a different proportion experiencing a toxicity event. As described in Appendix C, we traditionally apply a regression model when investigating the association between a binary response variable (denoted here by Y_i) and a binary treatment variable (given by T_i) of the form:

Equation 1

$$g(p_i) = g[P(Y_i = 1)] = \beta_0 + \underline{\beta}\underline{X}_i + \alpha T_i + \varepsilon_i$$

where

$g(.)$ = link function

α = parameter estimate of effect of interest

ε_i = random error term with zero mean

If all confounders are known and measured precisely through the covariate vector $\underline{X}_i$ then this model will give unbiased estimates of the treatment effect. However, if unmeasured confounding is present then Equation 1 will give biased estimates for the

parameter estimates contained in the vector $\underline{\beta}$ and the estimate of the treatment effect given by α.

Thus, after taking expectations, Equation 1 becomes:

Equation 1a:

$$E\{g(p_i) \mid \underline{X}, T_i, \text{sample selection rule}\} = \beta_0 + \underline{\beta}\underline{X}_i + \alpha T_i + E(\varepsilon_i \mid \text{sample selection rule})$$

The error term, ε_i, no longer has an expectation of zero, as it is now conditional on the sample selection rule. It follows that if we are able to correct for the effect of the unmeasured confounding on ε_i then it can be accounted for it the model. As a result, the error term will again have an expected value of zero and α will be an unbiased estimate of the treatment effect.

B.3.3 Standard generalised linear model

We begin by modelling T_i, the probability of receiving regimen A based on the observed covariates associated with treatment allocation that are given in the vector X^* (Equation 2). Note we are now making no assumptions about whether the random error term has zero expectation.

Equation 2:

$$Index = I = probit(\pi_i) = probit[P(T_i = 1)] = \beta_0^* + \underline{\beta}\underline{X}_i^* + \gamma_i$$

where

γ_i is a random error term

Now, assume without loss of generalisability that if the index, I, is greater than or equal to zero then the patient starts treatment regimen A, and if I is less than zero then the patient starts regimen B. Then patient *i* starts regimen A when:

Equation 3:

$$I = \beta_0^* + \underline{\beta}\underline{X}_i^* + \gamma_i > 0$$

$$\Rightarrow$$

$$\gamma_i > -(\beta_0^* + \underline{\beta}\underline{X}_i^*)$$

Thus, Equation 3 is equivalent to the sample selection rule described in Equation 1a. It is possible to show using standard bivariate normal distribution theory [396] that taking the expectation of $\varepsilon_i | \gamma_i > -(\beta_0^* + \underline{\beta}\underline{X}_i^*)$ leads to:

Equation 4:

$$E\left(\varepsilon_i | \gamma_i > -(\beta_0^* + \underline{\beta}\underline{X}_i^*)\right) = \frac{\text{cov}(\varepsilon_i, \gamma_i)}{\sqrt{\text{var}(\gamma_i)}} \lambda$$

$$= \tau_i \lambda_i$$

where

$$\lambda_i = \begin{cases} \dfrac{f[-(\beta_0^* + \beta^* X_i^*)]}{1 - F[-(\beta_0^* + \beta^* X_i^*)]} & \text{where } T_i = 1 \\ \dfrac{-f[-(\beta_0^* + \beta^* X_i^*)]}{F[-(\beta_0^* + \beta^* X_i^*)]} & \text{where } T_i = 0 \end{cases}$$

$f(.)$ is the value of the standard normal density function

$F(.)$ is the value of the cumulative normal density function

λ_i is the inverse mills ratio (IMR) for individual i.

Therefore, we now have an estimate, given by $\tau_i \lambda_i$, of the bias of ε_i. Combining the result of Equation 4 with Equation 1a leads to:

Equation 5:

$$probit(p_i) = probit[P(Y_i = 1)] = \beta_0 + \underline{\beta}\underline{X}_i + \alpha T_i + \tau_i \lambda_i + \varepsilon_i$$

where $E(\varepsilon_i) = 0$

As the error term now has an expectation of zero, carrying out a regression model which is adjusted for the IMR calculated from the first stage model will lead to unbiased estimates of α.

References

(1) Pneumocystis pneumonia -Los Angeles. Morbidity and Mortality Weekly Reports 1981; 30:250-252.

(2) Barre-Sinoussi F, Chermann JC, Rey F, Nugeyre MT, Chamaret S, Gruest J et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 1983; 220(4599):868-871.

(3) Tyrer F, Walker AS, Gillett J, Porter K. The relationship between HIV seroconversion illness, HIV test interval and time to AIDS in a seroconverter cohort. Epidemiol Infect 2003; 131(3):1117-1123.

(4) Bacchetti P, Moss AR. Incubation period of AIDS in San Francisco. Nature 1989; 338:251-253.

(5) Buchbinder SP, Katz MH, Hessol NA, O'Malley PM, Holmberg SD. Long-term HIV-1 infection without immunologic progression. AIDS 1994; 8:1123-1128.

(6) European Centre for the Epidemiological Monitoring of AIDS. 1993 revision of the European AIDS surveillance case definition. AIDS Surveillance in Europe, Quarterly Report 1993; 37:23-28.

(7) McCune JM. The dynamics of CD4+ T-cell depletion in HIV disease. Nature 2001; 410(6831):974-979.

(8) Weiss RA. Gulliver's travels in HIV land. Nature 2001; 410:963-967.

(9) Maini MK, Gilson RJC, Chavda N, Gill S, Fakoya A, Ross EJ et al. Reference ranges and sources of variability of CD4 counts in HIV-seronegative women and men. Genitourin Med 1996; 72:27-31.

(10) Bofill M, Janossy G, Lee CA, Macdonald-Burns D, Phillips AN, Sabin CA et al. Laboratory control values for CD4 and CD8 T lymphocytes. Implications for HIV-1 diagnosis. Clin Exp Immunol 1992; 88:243-252.

(11) Phillips AN, Lee CA, Elford J, Janossy G, Bofill M, Timms A et al. Prediction of progression to AIDS by analysis of CD4 lymphocyte counts in a haemophiliac cohort. AIDS 1989; 3:737-741.

(12) Mills GD, Jones PD. Relationship between CD4 lymphocyte count and AIDS mortality, 1986-1991. AIDS 1993; 7(10):1383-1386.

(13) Phillips AN, Lee CA, Elford J, Webster A, Janossy G, Timms A et al. More rapid progression to AIDS in older HIV-infected people: the role of CD4+ T-cell counts. J Acquir Immune Defic Syndr 1991; 4(10):970-975.

(14) Yarchoan R, Venzon DJ, Pluda JM, Lietzau J, Wyvill KM, Tsiatis AA et al. CD4 count and the risk for death in patients infected with HIV receiving antiretroviral therapy. Ann Intern Med 1991; 115(3):184-189.

(15) Koot M, Keet IP, Vos AH, de Goede RE, Roos MT, Coutinho RA et al. Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. Ann Intern Med 1993; 118(9):681-688.

(16) Schellekens PT, Tersmette M, Roos MT, Keet RP, de Wolf F, Coutinho RA et al. Biphasic rate of CD4+ cell count decline during progression to AIDS correlates with HIV-1 phenotype. AIDS 1992; 6(7):665-669.

(17) Margolick JB, Munoz A, Vlahov D, Astemborski J, Solomon L, He XY et al. Direct comparison of the relationship between clinical outcome and change in CD4+ lymphocytes in human immunodeficiency virus-positive homosexual men and injecting drug users. Arch Intern Med 1994; 154(8):869-875.

(18) Galai N, Munoz A, Chen K, Carey VJ, Chmiel J, Zhou SY. Tracking of markers and onset of disease among HIV-1 seroconverters. Stat Med 1993; 12(22):2133-2145.

(19) Phillips AN, Lee CA, Elford J, Janossy G, Kernoff PBA. The cumulative risk of AIDS as the CD4 lymphocyte count declines. Journal of Acquired Immune Deficiency Syndrome 1992; 5:148-152.

(20) Phillips AN, Sabin CA, Elford J, Bofill M, Janossy G, Lee CA. Acquired immunodeficiency syndrome (AIDS) risk in recent and long-standing human immunodeficiency virus type 1 (HIV-1)-infected patients with similar CD4 lymphocyte counts. Am J Epidemiol 1993; 138(10):870-878.

(21) Boutitie F, Pocock SJ. Predictive value of repeated measurements of CD4 lymphocyte counts on progression to AIDS. AIDS 1994; 8(1):35-41.

(22) Alcabes P, Selwyn PA, Davenny K, Hartel D, Buono D, Schoenbaum EE et al. Laboratory markers and the risk of developing HIV-1 disease among injecting drug users. AIDS 2003; 8(1):107-115.

(23) Phillips AN, Lee CA, Elford J, Janossy G, Timms A, Bofill M et al. Serial CD4 lymphocyte counts and development of AIDS. Lancet 1991; 337(8738):389-392.

(24) Post FA, Wood R, Maartens G. CD4 and total lymphocyte counts as predictors of HIV disease progression. QJM 1996; 89(7):505-508.

(25) Giorgi J, Lyles RH, Matud JL, Yamashita TE, Mellors JW, Hultin LE et al. Predictive value of immunologic and virologic markers after long or short duration of HIV-1 infection. J Acquir Immune Defic Syndr Hum Retrovirol 2002; 29(4):346-335.

(26) Mocroft A, Bofill M, Lipman M, Medina E, Borthwick N, Timms A et al. CD8+, CD38+ Lymphocyte percent: a useful immunological marker for monitoring HIV-1-infected patients. Journal of Acquired Immune Deficiency Syndrome 1997; 14(2):158-162.

(27) Burgisser P, Hammann C, Kaufmann D, Battegay M, Rutschmann OT. Expression of CD28 and CD38 by CD8+ T lymphocytes in HIV-1 infection correlates with markers of disease severity and changes towards normalization under treatment. The Swiss HIV Cohort Study. Clin Exp Immunol 1999; 115(3):458-463.

(28) Bisset LR, Cone RW, Fischer M, Battegay M, Vernazza PL, Dubs RW et al. Long-term evaluation of T-Cell subset changes after effective combination antiretroviral therapy during asymptomatic HIV-infection. Journal of Acquired Immune Deficiency Syndrome 2001; 27(3):266-271.

(29) Blankson JN, Gallant JE, Siliciano RF. Proliferative responses to human immunodeficiency virus type 1 (HIV-1) antigens in HIV-1-Infected patients with immune reconstitution. J Infect Dis 2001; 183:657-661.

(30) Angel JB, Parato KG, Kumar A, Kravcik S, Badley AD, Fex C et al. Progressive human immunodeficiency virus-specific immune recovery with prolonged viral suppression. J Infect Dis 2001; 183:546-554.

(31) Fauci AS. Multifactorial nature of human immunodeficiency virus disease: implications for therapy. Science 1993; 262(5136):1011-1018.

(32) Wei X, Ghosh SK, Taylor ME, Johnson VA, Emini EA, Deutsch P et al. Viral dynamics in human immunodeficiency virus type 1 infection. Nature 1995; 373(6510):117-122.

(33) Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature 1995; 373(6510):123-126.

(34) Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. Science 1996; 271(5255):1582-1586.

(35) Lyles RH, Graham NM, Astemborski J, Vlahov D, Margolick JB, Saah AJ et al. Cell-associated infectious HIV-1 viral load as a predictor of clinical progression and survival among HIV-1 infected injection drug users and homosexual men. Eur J Epidemiol 1999; 15(2):99-108.

(36) Carre N, Boufassa F, Hubert JB, Chavance M, Rouzioux C, Goujard C et al. Predictive value of viral load and other markers for progression to clinical AIDS after CD4+ cell count falls below 200/microL. SEROCO & HEMOCO Study Group. Int J Epidemiol 1998; 27(5):897-903.

(37) Mellors JW, Munoz A, Giorgi J, Margolis DM, Tassoni CJ, Gupta P et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Arch Intern Med 1997; 126(12):946-954.

(38) Mellors JW, Rinaldo C, Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. Science 1996; 272:1167-1170.

(39) Sabin C, Phillips A, Elford J, Griffiths P, Janossy G, Lee C. The progression of HIV disease in a haemophiliac cohort followed for 12 years. British Journal of Haematology 1993; 83:330-333.

(40) Darby SC, Rizza CR, Doll R, Spooner RJD, Stratton IM, Thakrar B. Incidence of AIDS and excess of mortality associated with HIV in haemophiliacs in the United Kingdon: report on behalf of the directors of haemophilia centres in the United Kingdom. Br Med J 1989; 298:1064-1068.

(41) Mientjes GHC, van Ameijden JC, van den Hoek JAR, Goudsmit J, Miedema F, Coutinho RA. Progression of HIV among injecting drug users: indications for a lower rate of progression among those who have frequently borrowed injecting equipment. AIDS 1993; 7:1363-1370.

(42) Carre N, Deveau C, Belanger F, Boufassa F, Persoz A, Jadand C. Effect of age and exposure group on the onset of AIDS in heterosexual and homosexual HIV-infected patients. AIDS 1994; 8:797-802.

(43) Fernandez-Cruz E, Desco M, Montes MG, Longo N, Gonzalez B, Zabay JM. Immunological and serological markers predictive of progression to AIDS in a cohort of HIV-infected drug users. AIDS 1990; 4:987-994.

(44) Coates RA, Farewell VT, Raboud J, Read SE, MacFadden DK, Calzavara LM. Cofactors of progression to acquired immunodeficiency syndrome in a cohort of male sexual contacts of men with human immunodeficiency virus disease. Am J Epidemiol 1990; 132:717-722.

(45) Munoz A, Vlahov D, Solomon L, Margolick JB, Bareta JC, Cohn S. Prognostic indicators for development of AIDS among intraveneous drug users. J Acquir Immune Defic Syndr 1992; 5:694-700.

(46) Sabin CA, Griffioen A, Yee TT, Emery VC, Herrero-Martinez E, Phillips AN et al. Markers of HIV-1 disease progression in individuals with haemophilia coinfected with hepatitus C virus: a longitudinal study. Lancet 2002; 360:1546-1551.

(47) Moyle G. Anaemia in persons with HIV infection: prognostic marker a contributor to morbidity. AIDS Rev 2002; 4:13-20.

(48) Moore RD, Keruly JC, Chaisson RE. Anaemia and survival in HIV infection. J Acquir Immune Defic Syndr 1998; 19:29-33.

(49) Semba RD, Shah N, Vlahov D. Improvement of anaemia among HIV-infected drug users receiving highly active antiretroviral therapy. J Acquir Immune Defic Syndr 2001; 26:315-319.

(50) Mocroft A, Kirk O, Barton SE, Lundgren JD. Anaemia is an independent predictive marker for clinical prognosis in HIV-infected patients from across Europe. AIDS 1999; 13:943-950.

(51) Medley GF, Anderson RM, Cox DR, Billard L. Incubation period of AIDS in patients infected via blood transfusion. Nature 1987; 328:719-721.

(52) Msellati P, Dupon M, Morlat P, Lacoste D, Pellegrin J, Dabis F. A cohort study of 89 HIV-1-infected adult patients contaminated by blood products: Bordeaux 1981-1989. AIDS 1990; 4:1105-1109.

(53) Friedland GH, Saltzman B, Vileno J, Freeman K, Schrager LK, Klein RS. Survival differences in patients with AIDS. J Acquir Immune Defic Syndr 1991; 4:144-153.

(54) Rothenberg R, Woelfel M, Stoneberger R, Millberg J, Parker R, Truman B. Survival with the acquired immunodeficiency syndrome. N Eng J Med 1987; 317:1297-1302.

(55) Melnick SL, Sherer R, Louis TA, Hillman D, Rodriguez EM, Lackman C. Survival and disease progression according to gender of patients with HIV infection. JAMA 1994; 272:1915-1921.

(56) Easterbrook PJ, Keruly JC, Creagh-Kirk T, Richman DD, Chaisson RE, Moore RD. Racial and ethnic differences in outcome in zidovudine-treated patients with advanced HIV disease. JAMA 1991; 266:2713-2718.

(57) Balotta C, Bagnarelli P, Violin M, Ridolfo AL, Zhou D, Berlusconi A et al. Homozygous delta 32 deletion of the CCR-5 chemokine receptor gene in an HIV-1-infected patient. AIDS 1997; 11(10):F67-F71.

(58) Biti R, Ffrench R, Young J, Bennetts B, Stewart G, Liang T. HIV-1 infection in an individual homozygous for the CCR5 deletion allele. Nat Med 1997; **3**(**3**):252-253.

(59) O'Brien TR, Winkler C, Dean M, Nelson JA, Carrington M, Michael NL et al. HIV-1 infection in a man homozygous for CCR5 delta 32. Lancet 1997; 349(9060):1219.

(60) Theodorou I, Meyer L, Magierowska M, Katlama C, Rouzioux C. HIV-1 infection in an individual homozygous for CCR5 delta 32. Seroco Stuy Group. Lancet 1997; 349(9060):1219-1220.

(61) Kageyama S, Mimaya J, Yamada K, Kurimura T, Shiraki K. Polymorphism of CCR5 affecting HIV disease progression in the Japanese population. AIDS Res Hum Retroviruses 2001; 17(11):991-995.

(62) Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 1996; 273(5283):1797-1798.

(63) Eugen-Olsen J, Iversen AK, Garred P, Koppelhus U, Pederson C, Benfield TL et al. Heterozygosity for a deletion in the CKR-5 gene leads to prolonged AIDS-free survival and slower CD4 T-cell decline in a cohort of HIV-seropositive individuals. AIDS 1997; 11(3):305-310.

(64) Michael NL, Chang G, Louie LG, Mascola JR, Dondero D, Birx DL et al. The role of viral phenotype and CCR-5 gene defects in HIV-1 transmission and disease progression. Nat Med 1997; **3**(**3**):338-340.

(65) Huang Y, Paxton WA, Wolinsky SM, Neumann AU, Zhang L, He T et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 1996; 2(11):1240-1243.

(66) Morawetz RA, Rizzardi GP, Glauser D, Rutschmann O, Hirschel B, Perrin L et al. Genetic polymorphism of CCR5 gene and HIV disease: the heterozygous (CCR5/delta ccr5) genotype is neither essential nor sufficient for protection against disease progression. Swiss HIV Cohort. Eur J Immunol 1997; 27(12):3223-3227.

(67) Mummidi S, Ahuja SS, Gonzalez E, Anderson SA, Santiago EN, Stephan KT et al. Genealogy of the CCR5 locus and chemokine system gene variants associated with altered rates of HIV-1 disease progression. Nat Med 1998; 4(7):786-793.

(68) Vella S, Floridia M. Saquinavir. Clinical pharmacology and efficacy. Clin Pharmacokinet 1998; 34(3):189-201.

(69) Figgitt DP, Plosker GL. Saquinavir soft-gel capsule: an updated review of its use in the management of HIV infection. Drugs 2000; 60(2):481-516.

(70) Smith CJ, Phillips AN, Hill T, Fisher M, Gazzard B, Porter K et al. The rate of viral rebound after attainment of viral load below 50 copies/ml according to specific antiretroviral drugs in use. J Infect Dis 2005; 192:1387-1397.

(71) Phillips AN, Ledergerber B, Horban A, Reiss P, Chiesi A, Kirk O et al. Rate of viral rebound according to specific drugs in the regimen in 2120 patients with HIV suppression. AIDS 2004; 18(13):1795-1804.

(72) Yu K, Daar ES. Dual protease inhibitor therapy in the management of the HIV-1. Expert Opin Pharmacother 2000; 1(7):1331-1342.

(73) Gazzard B, for the BHIVA writing committee. British HIV Association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy. HIV Med 2005; 6(S2):1-61.

(74) Nelson M, Fatkenheuer G, Konourina I, Lazzarin A, Clumeck N, Horban A et al. Efficacy and Safety of Maraviroc plus Optimized Background Therapy in Viremic, ART-experienced Patients Infected with CCR5-tropic HIV-1 in Europe, Australia, and North America: 24-Week Results. Fourteenth Conference on Retroviruses and Opportunistic Infections, Los Angeles 2007; 104aLB.

(75) Lalezari J, Goodrich J, DeJesus E, Lampiris H, Gulick R, Saag M et al. Efficacy and Safety of Maraviroc plus Optimized Background Therapy in Viremic ART-experienced Patients Infected with CCR5-tropic HIV-1: 24-Week Results of a Phase 2b/3 Study in the US and Canada. Fourteenth Conference on Retroviruses and Opportunistic Infections, Los Angeles 2007; 104bLB.

(76) Cooper D, Gatell J, Rockstroh J, Katlama C, Yeni P, Lazzarin A et al. Results of BENCHMRK-1, a Phase III Study Evaluating the Efficacy and Safety of MK-0518, a Novel HIV-1 Integrase Inhibitor, in Patients with Triple-class Resistant Virus. Fourteenth Conference on Retroviruses and Opportunistic Infections, Los Angeles 2007; 105aLB.

(77) Steigbigel R, Kumar P, Eron J, Schechter M, Markowitz M, Loufty M et al. Results of BENCHMRK-2, a Phase III Study Evaluating the Efficacy and Safety of MK-0518, a Novel HIV-1 Integrase Inhibitor, in Patients with Triple-class Resistant Virus. Fourteenth Conference on Retroviruses and Opportunistic Infections, Los Angeles 2007; 105bLB.

(78) Reddy S, Min S, Borland J, Song I, Lin J, Mehta A et al. A Double-blind, Parallel, Randomized, Placebo-controlled, Single and Repeat Dose-escalation Study to Investigate the Safety, Tolerability, and Pharmacokinetics of the HIV Integrase Inhibitor GSK364735 in Healthy Subjects (GRZ105655). Fourteenth Conference on Retroviruses and Opportunistic Infections, Los Angeles 2007; 562.

(79) Concorde: MRC/ANRS randomised double-blind controlled trial of immediate and deferred zidovudine in symptom-free HIV infection. Concorde Coordinating Committee. Lancet 1994; 343(8902):871-881.

(80) Delta: a randomised double-blind controlled trial comparing combinations of zidovudine plus didanosine or zalcitabine with zidovudine alone in HIV-infected individuals. Delta Coordinating Committee. Lancet 1996; 348(9023):283-291.

(81) Hammer SM, Katzenstein DA, Hughes MD, Gundacker H, Schooley RT, Haubrich RH et al. A Trial Comparing Nucleoside Monotherapy with Combination Therapy in HIV-Infected Adults with CD4 Cell Counts from 200 to 500 per Cubic Millimeter. N Eng J Med 1996; 335(15):1081-1090.

(82) Hammer SM, Katzenstein DA, Hughes MD, Gundacker H, Schooley RT, Haubrich RH et al. A trial comparing nucleoside monotherapy with combination therapy in HIV-infected adults with CD4 cell counts from 200 to 500 per cubic

millimeter. AIDS Clinical Trials Group Study 175 Study Team. N Eng J Med 1996; 335(15):1081-1090.

(83) Mauss S, Adams O, Willers R, Jablonowski H. Combination therapy with ZDV + DDI versus ZDV + DDC in patients with progression of HIV-infection under treatment with ZDV. J Acquir Immune Defic Syndr 1996; 11(5):469-477.

(84) Schooley RT, Ramirez-Ronda C, Lange JM, Cooper DA, Lavelle L, Lefkowitz L et al. Virologic and immunologic benefits of initial combination therapy with zidovudine and zalcitabine or didanosine compared with zidovudine monotherapy. Wellcome Resistance Study Collaborative Group. J Infect Dis 1996; 173(6):1354-1366.

(85) Hammer SM, Squires KE, Hughes MD, Grimes JM, Demeter LM, Currier JS et al. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. N Eng J Med 1997; 337(11):725-733.

(86) Palella FJ, Delaney KM, Moorman AC, Loveless MO, Fuhrer J, Satten GA et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. N Eng J Med 1998; 338(13):853-860.

(87) Mocroft A, Vella S, Benfield TL, Chiesi A, Miller V, Gargalianos P et al. Changing patterns of mortality across Europe in patients infected with HIV-1. Lancet 1998; 352:1725-1730.

(88) Mocroft A, Katlama C, Johnson AM, Pradier C, Antunes F, Mulcahy F et al. AIDS across Europe, 1994-98: the EuroSIDA study. Lancet 2000; 356:291-296.

(89) Egger M, Hirschel B, Francioli P, Sudre P, Wirz M, Flepp M et al. Impact of new antiretroviral combination therapies in HIV infected patients in Switzerland: prospective multicentre study. Br Med J 1997; 315:1194-1199.

(90) Olsen CH, Gatell J, Ledergerber B, Katlama C, Friis-Moller N, Weber J et al. Risk of AIDS and death at given HIV-RNA and CD4 cell counts, in relation to specific antiretroviral drugs in the regimen. AIDS 2005; 19(3):319-330.

(91) Lathey JL, Hughes MD, Fiscus SA, Pi T, Jackson JB, Rasheed S et al. Variability and prognostic values of virologic and CD4 cell measures in human immunodeficiency virus type 1-infected patients with 200-500 CD4 cells/mm(3) (ACTG 175). AIDS Clinical Trials Group Protocol 175 Team. J Infect Dis 1998; 177(3):617-24.

(92) O'Brien WA, Hartigan PM, Martin D, Esinhart J, Hill A, Benoit S et al. Changes in Plasma HIV-1 RNA and CD4+ Lymphocyte Counts and the Risk of Progression to AIDS. N Eng J Med 1996; 334(7):426-431.

(93) O'Brien WA, Hartigan PM, Daar ES, Simberkoff MS, Hammann C. Changes in Plasma HIV RNA Levels and CD4+ Lymphocyte Counts Predict Both Response to Antiretroviral Therapy and Therapeutic Failure. Annals of Internal Medicine 1997; 126(12):939-945.

(94) Ledergerber B, Lundgren JD, Walker AS, Sabin C, Justice A, Reiss P et al. Predictors of trend in CD4-positive T-cell count and mortality among HIV-1-infected individuals with virological failure to all three antiretroviral-drug classes. Lancet 2004; 364(9428):51-62.

(95) CASCADE Collaboration. Determinants of survival following HIV-1 seroconversion after the introduction of HAART. Lancet 2003; 362:1267-1274.

(96) Connick E, Lederman MM, Kotzin BL, Spritzler J, Kuritzkes DR, St Clair M et al. Immune reconstitution in the first year of potent antiretroviral therapy and its relationship to virologic response. J Infect Dis 2000; 181:358-363.

(97) Smith CJ, Sabin CA, Youle MS, Kinloch-de-Loes S, Lampe FC, Madge S et al. Factors influencing increases in CD4 cell counts of HIV-positive persons receiving long-term highly active antiretroviral therapy. J Infect Dis 2004; **190(10)**:1860-1868.

(98) Staszewski S, Miller V, Sabin CA, Schlecht C, Gute P, Stamm S et al. Determinants of sustainable CD4 lymphocyte count increases in response to antiretroviral therapy. AIDS 1999; 13(8):951-956.

(99) Hejdeman B, Lenkel R, Leandersson AC, Hultstrom AL, Wahren B, Sandstrom E et al. Clinical and immunological benefits from highly active antiretroviral therapy in spite of limited viral load reduction in HIV Type 1 infection. AIDS Res Hum Retroviruses 2001; 17(4):277-286.

(100) Li TS, Tubiana R, Katlama C, Calvez V, Ait Mohand H, Autran B. Long-lasting recovery in CD4 T-cell function and viral-load reduction after highly active antiretroviral therapy in advanced HIV-1 disease. Lancet 1998; 351:1682-1686.

(101) Brun-Vezinet F, Boucher C, Loveday C, Descamps D, Fauveau V, Izopet J et al. HIV-1 viral load, phenotype, and resistance in a subset of drug-naive participants from the Delta trial. Lancet 1997; 350(9083):983-990.

(102) Katzenstein DA, Hammer SM, Hughes MD, Gundacker H, Jackson JB, Fiscus S et al. The Relation of Virologic and Immunologic Markers to Clinical Outcomes after Nucleoside Therapy in HIV-Infected Adults with 200 to 500 CD4 Cells per Cubic Millimeter. N Eng J Med 1996; **335(15)**:1091-1098.

(103) Kaufmann GR, Khanna N, Weber R, Perrin L, Furrer H, Cavassini M et al. Long-term virological response to multiple sequential regimens of highly active antiretroviral therapy for HIV infection. Antivir Ther 2004; **9(2)**:263-274.

(104) Parruti G, Manzoli L, Toro PM, D'amico G, Rotolo S, Graziani V et al. Long-term adherence to first-line highly active antiretroviral therapy in a hospital-based cohort: predictors and impact on virologic response and relapse. AIDS Patient Care and STDs 2006; 20(1):48-56.

(105) Parienti JJ, Massari V, Descamps D, Vabret A, Bouvet E, Larouze B et al. Predictors of virologic failure and resistance in HIV-infected patients treated with nevirapine- or efavirenz-based antiretroviral therapy. Clin Infect Dis 2004; **38**:1311-1316.

(106) Fletcher CV, Testa MA, Brundage RC, Chesney MA, Haubrich R, Acosta EP et al. Four measures of antiretroviral medication adherence and virologic response in AIDS clinical trials group study 359. J Acquir Immune Defic Syndr 2005; 40(3):301-306.

(107) Nieuwkerk PT, Oort FJ. Self-reported adherence to antiretroviral therapy for HIV-1 infection and virologic treatment response: a meta-analysis. J Acquir Immune Defic Syndr 2005; 38(4):445-448.

(108) Press N, Tyndall MW, Wood E, Hogg RS, Montaner JS. Virologic and immunologic response, clinical progression, and highly active antiretroviral therapy adherence. J Acquir Immune Defic Syndr 2002; 31(S3):S112-S117.

(109) Haubrich RH, Little SJ, Currier JS, Forthal DN, Kemper CA, Beall GN et al. The value of patient-reported adherence to antiretroviral therapy in predicting virologic and immunologic response. California Collaborative Treatment Group. AIDS 1999; 13(9):1099-1107.

(110) Escobar I, Campo M, Martin J, Fernandez-Shaw C, Pulido F, Rubio R. Factors affecting patient adherence to highly active antiretroviral therapy. Ann Pharmacother 2003; 37(6):775-781.

(111) Hinkin CH, Hardy DJ, Mason KI, Castellon SA, Durvasula RS, Lam MN et al. Medication adherence in HIV-infected adults: Effect of patient age, cognitive status, and substance abuse. AIDS 2004; 18(S1):S19-S25.

(112) Turner BJ, Laine C, Cosler L, Hauck WW. Relationship of gender, depression, and health care delivery with antiretroviral adherence in HIV-infected drug users. J Gen Intern Med 2003; 18(4):248-257.

(113) Kuyper LM, Wood E, Montaner JSG, Yip B, O'Connell JM, Hogg RS. Gender differences in HIV-1 RNA rebound attributed to incomplete antiretroviral adherence among HIV-infected patients in a population-based cohort. J Acquir Immune Defic Syndr 2004; 37(4):1470-1476.

(114) Berg KM, Demas PA, Howard AA, Schoenbaum EE, Gourevitch MN, Arnsten JH. Gender differences in factors associated with adherence to antiretroviral therapy. J Gen Intern Med 2004; 19(11):1111-1117.

(115) Frater AJ, Dunn DT, Beardall AJ, Ariyoshi K, Clarke JR, McClure MO et al. Comparative response of African HIV-1-infected individuals to highly active antiretroviral therapy. AIDS 2002; 16(8):1139-1146.

(116) Anastos K, Schneider MF, Gange SJ, Minkoff H, Greenblatt RM, Feldman J et al. The association of race, sociodemographic, and behavioral characteristics with response to highly active antiretroviral therapy in women. J Acquir Immune Defic Syndr 2005; 39(5):537-544.

(117) Bangsberg DR. Less than 95% adherence to nonnucleoside reverse-transcriptase inhibitor therapy can lead to viral suppression. Clin Infect Dis 2006; 43(7):939-941.

(118) Egger M, May M, Chene G, Phillips AN, Ledergerber B, Dabis F et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. Lancet 2002; 360:119-129.

(119) Lucas GM, Chaisson RE, Moore RD. Highly active antiretroviral therapy in a large urban clinic: risk factors for virologic failure and adverse drug reactions. Ann Intern Med 1999; 131(2):81-87.

(120) Staszewski S, Miller V, Sabin C, Carlebach A, Berger AM, Weidmann E et al. Virological response to protease inhibitor therapy in an HIV clinic cohort. AIDS 1999; 13(3):367-373.

(121) Mocroft A, Gill MJ, Davidson W, Phillips AN. Predictors of a viral response and subsequent virological treatment failure in patients with HIV starting a protease inhibitor. AIDS 1998; 12(16):2161-2167.

(122) Skowren G, Street JC, Obee EM. Baseline CD4 + Cell Count, Not Viral Load, Correlates With Virologic Suppression Induced by Potent Antiretroviral Therapy. J Acquir Immune Defic Syndr 2001; 28(4):313-319.

(123) Phillips AN, Staszewski S, Lampe F, Youle M, Klauke S, Bickel M et al. Human immunodeficiency virus rebound after suppression to <400 copies/mL during initial highly active antiretroviral therapy regimens, according to prior nucleoside experience and duration of suppression. J Infect Dis 2002; 186(8):1086-1091.

(124) Dornadula G, Zhang H, VanUitert B, Stern J, Livornese Jr L, Ingerman MJ et al. Residual HIV-1 RNA in blood plasma of patients taking suppressive highly active antiretroviral therapy. JAMA 1999; 282:1627-1632.

(125) Furtaldo MR, Callaway DS, Phair JP, Kunstman BS, Stanton JL, Macken CA et al. Persistence of HIV-1 transcription in peripheral blood mononuclear cells in patients receiving potent antiretroviral therapy. N Eng J Med 1999; 340:1614-1622.

(126) Sharkey ME, Teo I, Greenough T, Sharova N, Luzugisga K, Sullivan JL et al. Persistence of episomal HIV-1 infection intermediates in patients on highly active anti-retroviral therapy. Nat Med 2000; 6(1):76-81.

(127) Zhang L, Ramratnam B, Tenner-Racz K, He Y, Vesanen M, Lewin S et al. Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. N Eng J Med 1999; 340(21):1605-1613.

(128) Wong JK, Hezareh M, Gunthard HF, Havlir DV, Ignacio CC, Spina CA et al. Recovery of replication-competent HIV despite prolonged suppression of plasma viremia. Science 1997; 278(5341):1291-1295.

(129) Finzi D, Hermankova M, Pierson T, Carruth LM, Buck C, Chaisson RE et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. Science 1997; 278(5341):1295-300.

(130) Chun TW, Stuyver L, Mizell SB, Ehler LA, Mican JA, Baseler M et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. Proc Natl Acad Sci U S A 1997; 94(24):13193-13197.

(131) Kaufmann GR, Bloch M, Zaunders JJ, Smith D, Cooper DA. Long-term immunological response in HIV-1-infected subjects receiving potent antiretroviral therapy. AIDS 2000; 14:959-969.

(132) Smith CJ, Sabin CA, Lampe F, Kinloch-de-Loes S, Gumley H, Carroll A et al. The potential for CD4 cell increases in HIV-positive individuals who control viraemia with highly active antiretroviral therapy (HAART). AIDS 2003; 17(7):963-969.

(133) Hirsch MS, Conway B, D'Aquila RT, Johnson VA, Brun-Vezinet F, Clotet B et al. Antiretroviral drug resistance testing in adults with HIV infection: Implications for clinical management. JAMA 1998; 279(24):1984-1991.

(134) Johnson VA, Brun-Vezinet F, Clotet B, Conway B, Kuritzkes DR, Pillay D et al. Update of the drug resistance mutations in HIV-1:2005. Topics in HIV Medicine International AIDS Society-USA 2005; 13(1):51-57.

(135) Phillips AN, Dunn D, Sabin C, Pozniak A, Matthias R, Geretti AM et al. Long term probability of detection of HIV-1 drug resistance after starting antiretroviral therapy in routine clinical practice. AIDS 2005; 19(5):487-494.

(136) Cane P, Chrystie I, Dunn D, Evans B, Geretti AM, Green H et al. Time trends in primary resistance to HIV drugs in the United Kingdom: multicentre observational study. Br Med J 2005; 332(7534):179-180.

(137) Bartlett JA. Addressing the Challenges of Adherence. J Acquir Immune Defic Syndr 2002; 29(S1):S2-S10.

(138) Moyle G, DeJesus E, Cahn P, Castillo SA, Zhao H, Gordon DN et al. Abacavir Once or Twice Daily Combined With Once-Daily Lamivudine and Efavirenz for the Treatment of Antiretroviral-Naive HIV-Infected Adults: Results of the Ziagen Once Daily in Antiretroviral Combination Study. J Acquir Immune Defic Syndr 2005; 38(4):417-425.

(139) DeJesus E, Herrera G, Teofilo E. Abacavir versus zidovudine combined with lamivudine and efavirenz, for the treatment of antiretroviral naive HIV-infected individuals. Clin Infect Dis 2004; 39:1038-1046.

(140) Fortin C, Joly V. Efavirenz for HIV-1 infection in adults: An overview. Expert Rev Anti Inf 2004; 2(5):671-684.

(141) van Leth F, Andrews S, Grinsztejn B, Wilkins E, Lazanas MK, Lange JMA et al. The effect of baseline CD4 cell count and HIV-1 viral load on the efficacy and safety of nevirapine or efavirenz-based first-line HAART. AIDS 2005; **19(5)**:463-471.

(142) Barrios A, Rendon AL, Gallego O, Martin-Carbonero L, Valer L, Rios P et al. Predictors of virological response to atazanavir in protease inhibitor-experienced patients. HIV Clin Trials 2004; 5(4):201-205.

(143) White AJ. Mitochondrial toxicity and HIV therapy. Sex Transm Infect 2001; **77(3)**:158-173.

(144) Semba RD, Shah N, Vlahov D. Risk factors and cumulative incidence of anaemia among HIV-infected injection drug users. Int J STD AIDS 2002; **13(2)**:119-123.

(145) Lalezari JP, Henry K, O'Hearn M, Montaner JSG, Piliero PJ, Trottier B et al. Enfuvirtide, an HIV-1 Fusion Inhibitor, for Drug-Resistant HIV Infection in North and South America. N Eng J Med 2003; 348(22):2175-2185.

(146) Lazzarin A, Clotet B, Cooper D, Reynes J, Arasteh K, Nelson M et al. Efficacy of Enfuvirtide in Patients Infected with Drug-Resistant HIV-1 in Europe and Australia. N Eng J Med 2003; 348(22):2186-2195.

(147) DeCock KM, Glenn Fowler M, Mercier E, de Vincenzi I, Saba J, Hoff E et al. Prevention of mother-to-child HIV transmission in resource-poor countries. J Am Med Assoc 2000; 283(9):1175-1182.

(148) Kind C, Rudin C, Siegrist CA, Wyler CA, Bidermann K, Lauper U et al. Prevention of vertical HIV transmission: additive protective effect of elective Cesarean section and zidovudine prophylaxis. AIDS 1998; 12(2):205-210.

(149) Mandelbrot L, Le Chenadec J, Berrebi A, Bongain A, Benifla JL, Delfraissy JF et al. Perinatal HIV-1 Transmission; Interaction Between Zidovudine Prophylaxis and Mode of Delivery in the French Perinatal Cohort . J Am Med Assoc 1998; 280(1):55-60.

(150) Pliner V, Weedon J, Thomas P, Steketee RW, Abrams EJ, Lambert G et al. Incubation period of HIV-1 in perinatally infected children. AIDS 1998; 12(7):759-766.

(151) Blanche S, Newell ML, Mayaux MJ, Dunn DT, Teglas JP, Rouzioux C et al. Morbidity and mortality in European children vertically infected by HIV-1. The French Pediatric HIV Infection Study Group and European Collaborative Study. J Acquir Immune Defic Syndr 1997; 14(5):442-450.

(152) Bamji M, Thea DM, Weedon J, Krasinski K, Matheson PB, Thomas P et al. Prospective study of human immunodeficiency virus 1-related disease among 512 infants born to infected women in New York City. The New York City Perinatal HIV Transmission Collaborative Study Group. Pediatr Infect Dis J 1996; 15(10):891-898.

(153) Tovo PA, de Martino M, Gabiano C, Cappello N, D'Elia R, Loy A et al. Prognostic factors and survival in children with perinatal HIV-1 infection. The Italian Register for HIV Infections in Children. Lancet 1992; 339(8804):1249-1253.

(154) European Centre for the Epidemiological Monitoring of AIDS. European case definition for AIDS surveillance in children - revision 1995. HIV/AIDS Surveillance in Europe, Quarterly Report 1995; 48:46-53.

(155) Sharland M, Blanche S, Castelli G, Ramos J, Gibb DM. PENTA guidelines for the use of antiretroviral therapy, 2004. HIV Med 2004; 5(S2):61-86.

(156) Mofensen LM, Korelitz J, Meyer WA, Bethel J, Rich K, Pahwa S et al. The relationship between serum human immunodeficiency virus type 1 (HIV-1) RNA level, CD4 lymphocyte percent, and long-term mortality risk in HIV-infected children. J Infect Dis 1997; 175:1029-1038.

(157) HIV Paediatric Prognostic Markers Collaborative Study Group. Short-term risk of disease progression in HIV-1-infected children receiving no antiretroviral therapy or zidovudine monotherapy: a meta-analysis. Lancet 2003; 362(9396):1605-1611.

(158) Paediatric European Network for Treatment of AIDS (PENTA). HIV-1 viral load and CD4 cell count in untreated children with vertically acquired asymptomatic or mild disease. Paediatric European Network for Treatment of AIDS (PENTA). AIDS 1998; 12(4):F1-F8.

(159) Shearer WT, Quinn TC, LaRussa P, Lew JF, Mofenson L, Almy S et al. Viral Load and Disease Progression in Infants Infected with Human Immunodeficiency Virus Type 1. N Eng J Med 1997; 336(19):1337-1342.

(160) McIntosh K, Shevitz A, Zaknun D, Kornegay J, Chatis P, Karthas N et al. Age- and Time-related Changes in Extracellular Viral Load in Children Vertically Infected by Human Immunodeficiency Virus. Pediatr Infect Dis J 1996; 15(12):1087-1091.

(161) De Rossi A, Masiero S, Giaquinto C, Ruga E, Comar M, Giacca M et al. Dynamics of Viral Replication in Infants with Vertically Acquired Human Immunodeficiency Virus Type 1 Infection. J Clin Invest 1996; 97(2):323-330.

(162) O'Brien T.R., Blattner WA, Waters D, Eyster E, Hilgartner MW, Cohen AR et al. Serum HIV-1 RNA levels and time to development of AIDS in the Multicenter Hemophilia Cohort Study. JAMA 1996; 276(2):105-110.

(163) Resion S, Bellon J, Gurbindo D, Tomas Ramos J, Antonio Leon J, Lose Mellado M et al. Viral load and CD4+ T lymphocyte response to highly active antiretroviral therapy in human immunodeficiency virus type 1-infected children: an observational study. Clin Infect Dis 2003; 37(9):1216-1225.

(164) Kovacs A, Montepiedra G, Carey V, Pahwa S, Weinberg A, Frenkel L et al. Immune reconstitution after receipt of highly active antiretroviral therapy in children with advanced or progressive HIV disease and complete or partial viral load response. J Infect Dis 2005; 192(2):296-302.

(165) Rosenblatt HM, Stanley KE, Song LY, Johnson GM, Wiznia AA, Nachman SA et al. Immunological response to highly active antiretroviral therapy in children with clinically stable HIV-1 infection. J Infect Dis 2005; 192(3):445-455.

(166) Sabin CA, for the Collaboration of Observational HIV Epidemiological Research Europe (COHERE) Study. Response to Combination Antiretroviral Therapy (cART): Variation by Age. Fourteenth Conference on Retroviruses and Opportunistic Infections, Los Angeles 2007; Abstract 528.

(167) Leonard EG, McComsey GA. Metabolic complications of antiretroviral therapy in children. Pediatr Infect Dis J 2003; 22(1):77-84.

(168) Church JA, Mitchell WG, Gonzalez-Gomez I, Christensen J, Vu TH, Dimauro S et al. Mitochondrial DNA depletion, near-fatal metabolic acidosis, and liver failure in an HIV-infected child treated with combination antiretroviral therapy. J Pediatr 2001; 138(5):748-751.

(169) Mocroft A, Ledergerber B, Katlama C, Kirk KA, Reiss P, d'Arminio Monforte A et al. Decline in the AIDS and death rates in the EuroSIDA study: an observational study. Lancet 2003; 362:22-29.

(170) Powderly WG. Long-term exposure to lifelong therapies. Journal of Acquired Immune Deficiency Syndrome 2002; 29:S28-S40.

(171) Arenas-Pinto A, Grant AD, Edwards S, Weller IVD. Lactic acidosis in HIV infected patients: a systematic review of published cases. Sex Transm Infectc 2003; 79(340):344.

(172) Mondy K, Yarasheski K, Powderly WG, Whyte M, Claxton S, DeMarco D et al. Longitudinal evolution of bone mineral density and bone markers in human immunodeficiency virus-infected individuals. Clin Infect Dis 2003; 36:482-490.

(173) Carr A, Cooper DA. Adverse effects of antiretroviral therapy. Lancet 2000; 356:1423-1430.

(174) Hewitt RG. Abacavir hypersensitivity reaction. Clin Infect Dis 2002; 15:1137-1142.

(175) Fagot JP, Mockenhaupt M, Bouwes-Bavinck JN, Naldi L, Viboud C, Roujeau JC. Nevirapine and the risk of Stevens-Johnson syndrome or toxic epidermal necrolysis. AIDS 2001; 15(1843):1848.

(176) Sundar K. Zidovudine-induced fatal lactic acidosis and hepatic failure in patients with acquired immunodeficiency syndrome: report of two patients and review of the literature. Crit Care Med 1997; 8:1425-1430.

(177) Sabin CA. The role of observational studies in assessing the impact of antiviral therapies. Curr Opin Infect Dis 2000; 13(6):631-635.

(178) Dunn D, Babiker AG, Hooker M, Darbyshire JH. The dangers of inferring treatment effects from observational data: a case study in HIV infection. Control Clin Trials 2002; 23:106-110.

(179) Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the heirarchy of research designs. N Eng J Med 2000; 342(25):1887-1892.

(180) Joly V, Flandre P, Meiffredy V, Leturque N, Harel M, Aboulker JP et al. Increased risk of lipoatrophy under stavudine in HIV-1-infected patients: results of a substudy from a comparative trial. AIDS 2002; 16(18):2447-2454.

(181) van Leth F, Phanuphak P, Gazzard B, Cahn P, Wood R, Bloch M et al. Lipid changes in a randomised comparative trial of first-line antiretrovrial therapy with regimens containing either nevirapine alone, efavirinz alone or both drug combined, together with stavudine and lamivudine (2NN study). 10th Conference on retroviruses and opportunistic infections, Boston, USA 2003;Abstract 752.

(182) Matheron S, Massip P, Trepo C, Lepeu G, Delfraissy JF, Gastaut JA et al. Metabolic and clinical evaluation of lipodystrophy syndrome (LD) in HIV-1-infected adults receiving initial HAART with or without a protease inhibitor (PI): 48 week data fro the CNAF3007 study. 8th Conference on Retroviruses and Opportunistic Infections 2001;Abstract 670.

(183) van der Valk M, Gisolf EH, Wit FW, Japour A, Weverling GJ, Danner SA. Increased risk of lipodystrophy when nucleoside analogue reverse transcriptase inhibitors are included with protease inhibitors in the treatment of HIV-1 infection. AIDS 2001; 15(7):847-855.

(184) Podzamczer D, Ferrer E, Consiglio E, Gatell J, Perez P, Perez JL et al. A randomised clinical trial comparing nelfinavir or nevirapine associated to zidovudine/lamivudine in HIV-infected naive patients (the Combine study). Antivir Ther 2002; 7:81-90.

(185) Ungsedhapand C, Kroon EDMB, Suwanagool S, Ruxrungtham K, Yimsuan N, Sonjai A et al. A randomised, open-label, comparative trial of zidovudine plus lamivudine vs zidovudine plus lamivudine plus didanosine in antiretroviral naive HIV-1-infected Thai patients. J Acquir Immune Defic Syndr 2001; 27(2):116-123.

(186) van Leeuwen R, Katlama C, Murphy RL, Squires K, Gatell J, Horban A et al. A randomized trial to study first-line combination therapy with or without a protease inhibitor in HIV-1-infected patients. AIDS 2003; 17:987-999.

(187) Moore DA, Goodall RL, Ives NJ, Hooker M, Gazzard BG, Easterbrook PJ. How generalizable are the results of large randomized controlled trials of antiretroviral therapy? HIV Med 2000; 1(3):149-154.

(188) Hankins C, Lapointe N, Walmsley S. Participation in clinical trials among women living with HIV in Canada. Canadian Women's HIV Study Group. CMAJ 1998; 159(11):1359-1354.

(189) Madge S, Mocroft A, Wilson D, Youle M, Phillips AN, Tyrer M et al. Participation in clinical studies among patients infected with HIV-1 in a single treatment centre over 12 years. HIV Med 2000; 1:212-218.

(190) Gifford AL, Cunningham WE, Heslin KC, Andersen RM, Nakazono T, Lieu DK et al. Participation in research and access to experimental treatments by HIV-infected patients. N Eng J Med 2002; 346(18):1373-1382.

(191) Lederman MM, Connick E, Landay A, Kuritzkes DR, Spritzler J, St Clair M et al. Immunological responses associated with 12 weeks of combination antiretroviral therapy consisting of Zidovudine, Lamivudine, and Ritonavir: Results of AIDS Clinical Trials Group Protocol 315. J Infect Dis 1998; 178:70-79.

(192) Walmsley S, Bernstein B, King M, Arribas J, Beall G, Ruane P et al. Lopinavir-Ritonavir versus Nelfinavir for the initial treatment of HIV infection. N Eng J Med 2002; 346(26):2039-2046.

(193) Staszewski S, Keiser P, Montaner J, Raffi F, Gathe J, Brotas V et al. Abacavir-Lamivudine-Zidovudine vs Indinavir-Lamivudine-Zidovudine in antiretroviral naive HIV-infected patients: A randomised equivalence trial. JAMA 2001; 285(9):1155-1163.

(194) Arnaiz JA, Mallolas J, Podzamczer D, Gerstoft J, Lundgren JD, Cahn P et al. Continued indinavir versus switching to indinavir/ritonavir in HIV-infected patients with suppressed viral load. AIDS 2003; 17(6):831-840.

(195) The Data Collection on Adverse Events of Anti-HIV Drugs (DAD) Study Group. Combination antiretroviral therapy and the risk of myocardial infarction. N Eng J Med 2003; 349(21):1993-2003.

(196) Phillips A, Carr A, Neuhaus J, Visnegarwala F, Prineas R, Burman W et al. Interruption of ART and Risk of Cardiovascular Disease: Findings from SMART. 14th Conference on Retrovirus and Opportunistic Infections 2007; O41.

(197) Hughes MD, Williams PL. Challenges in Using Observational Studies to Evaluate Adverse Effects of Treatment. N Eng J Med 2007; 356(17):1705-1707.

(198) Padberg J, Fritsche L, Bergmann F, Schurmann D, Suttorp N. Nephropathy and renal colic in patients treated with indinavir, ritonavir plus indinavir or ritonavir plus saquinavir. AIDS 1999; 13(15):2173.

(199) Herman JS, Ives NJ, Nelson M, Gazzard BG, Easterbrook PJ. Incidence and risk factors for the development of indinavir-associated renal complications. J Antimicrob Chemother 2001; 48:355-360.

(200) Boubaker K, Sudre P, Bally F, Vogel G, Meuwly JY, Glauser MP et al. Changes in renal function associated with indinavir. Actas Urol Esp 1998; 12(18):F249-F254.

(201) Klein D, Hurley LB, Quesenberry CP, Sidney S. Do protease inhibitors increase the risk for coronary heart disease in patients with HIV-1 infection? J Acquir Immune Defic Syndr Hum Retrovirol 2002; 30:471-477.

(202) Keys AB. Seven Countries: a multivariate analysis of death and coronary heart disease. Cambridge, Mass: Harvard University Press; 1980.

(203) Coresh J, Toto RD, Kirk KA, Whelton PK, Massry S, Jones C et al. Creatinine clearance as a measure of GFR in screenees for the African-American Study of Kidney Disease and Hypertension pilot study. Am J Kidney Dis 1998; 32(1):32-42.

(204) Toto RD, Kirk KA, Coresh J, Jones C, Appel L, Wright J et al. Evaluation of serum creatinine for estimating glomerular filtration rate in African Americans with hypertensive nephrosclerosis: results from the African-American Study of Kidney Disease and Hypertension (AASK) Pilot Study. J Am Soc Nephrol 1997; 8(2):279-287.

(205) Friss-Moller N, Weber R, Reiss P, Thiebaut R, Kirk O, d'Arminio Monforte A et al. Cardiovascular disease risk factors in HIV patients-association with antiretroviral therapy. AIDS 2003; 17(8):1179-1193.

(206) Hadigan C, Meigs JB, Corcoran C, Reitschel P, Basgoz N, Davis B et al. Metabolic abnormalities and cardiovascular disease risk factors in adults with human immunodeficiency virus infection and lipodystrophy. Clin Infect Dis 2001; 32:130-139.

(207) Carr A, Samaras K, Thorisdottir A, Kaufmann GR, Chisholm DJ, Cooper DA. Diagnosis, prediction and natural course of HIV-1 protease inhibitor-associated lipodystrophy, hyperlipidaemia and diabetes mellitus. Lancet 1999; 353:2093-2099.

(208) Behrens G, Dejam A, Schmidt H, Balks HJ, Brabant G, Korner T et al. Impaired glucose tolerance, beta cell function and lipid metabolism in HIV patients under treatment with protease inhibitors. AIDS 1999; 13(10):F63-F70.

(209) Collazos J, Mayo J, Ibarra S, Cazallas J. Hyperlipidemia in HIV-infected patients: the protective effect of hepatitis C virus co-infection. AIDS 2003; 17(6):927-929.

(210) Martinez E, Conget I, Lozano L, Casamitjana R, Gattel J. Reversion of metabolic abnormalities after switching from HIV-1 protease inhibitors to nevirapine. AIDS 1999; 13(7):805-810.

(211) Negredo E, Ribalta J, Parades R, Ferre R, Sirera G, Ruiz L et al. Reversal of atherogenic lipoprotein profile in HIV-1 infected patients with lipodystrophy after replacing protease inhibitors by nevirapine. AIDS 2002; 16(10):1383-1389.

(212) Moyle G, Baldwin C, Langroudi B, Mandalia S, Gazzard BG. A 48-week, randomized, open-label comparison of three abacavir-based substitution approaches in the management of dyslipidemia and peripheral lipoatrophy. J Acquir Immune Defic Syndr 2003; 33(1):22-28.

(213) Heath KV, Hogg R, Singer J, Chan KJ, O'Shaughnessy MV, Montaner JSG. Antiretroviral treatment patterns and incident HIV-associated morphologic and lipid abnormalities in a population-based cohort. J Acquir Immune Defic Syndr Hum Retrovirol 2002; 30(4):440-447.

(214) Dong KL, Bausserman LL, Flynn MM, Dickinson BP, Flanigan TP, Mileno MD et al. Changes in body habitus and serum lipid abnormalities in HIV-positive women on highly active antiretroviral therapy (HAART). J Acquir Immune Defic Syndr Hum Retrovirol 1999; 21(2):107-113.

(215) Rakotoambinina B, Medioni J, Rabian C, Jubault V, Jais J-P, Viard JP. Lipodystrophic syndromes and hyperlipidemia in a cohort of HIV-1-infected patients receiving triple combination antiretroviral therapy with a protease inhibitor. J Acquir Immune Defic Syndr Hum Retrovirol 2001; 27:443-449.

(216) Costa A, Pulido F, Rubio R, Cepeda C, Torralba M, Costa JR. Lipid changes in HIV-infected patients who started rescue therapy with an amprenavir/ritonavir-based highly active antiretroviral therapy. AIDS 2002; 16(14):1983-1984.

(217) Periard D, Telenti A, Sudre P, Cheseaux JJ, Halfon P, Reymond M et al. Atherogenic dyslipidemia in HIV-infected individuals treated with Protease Inhibitors. Circulation 1999; 100(7):700-705.

(218) Tsiodras S, Mantzoros C, Hammer S, Samore M. Effects of protease inhibitors on hyperglycemia, hyperlipiemia and lipodystrophy. Arch Intern Med 2000; 160(13):2050-2056.

(219) John M, McKinnon EJ, James IA, Nolan DA, Herrmann SA, Moore CB et al. Randomized, controlled, 48-week study of switching stavudine and/or protease inhibitors to combivir/abacavir to prevent of reverse lipoatrophy in HIV-infected patients. J Acquir Immune Defic Syndr 2003; 33(1):29-33.

(220) Hatuno H, Miller KD, Yoder CP, Yanovski JA, Sebring NG, Jones EC et al. Metabolic and anthropometric consequences of interruption of highly active antiretroviral therapy. AIDS 2000; 14:1935-1942.

(221) Curtis H, Sabin CA, Johnson MA. Findings from the first national clinical audit of treatment for people with HIV. HIV Med 2003; 4(1):11-17.

(222) Sabin CA, Phillips AN. Treatment comparisons in HIV infection: the benefits and limitations of observational cohort studies. J Antimicrob Chemother 2001; 47:371-375.

(223) Piketty C, Peytavin G, Trylezinski A, Gonzalez-Canali G, Castiel P, Weiss L et al. Low increase in serum lipids in patients receiving a combination of ritonavir, saquinavir and efavirinz. HIV Med 2002; 3(4):287-289.

(224) Paparizos VA, Kyriakis KP, Botsis C, Papastamopoulos V, Hadjivassiliou M, Stavrianeas NG. Protease inhibitor therapy-associated lipodystrophy, hypertriglyceridaemia and diabetes mellitus. AIDS 2000; 14(7):903.

(225) Carr A, Samaras K, Burton S, Law M, Freund J, Chisholm DJ et al. A syndrome of peripheral lipodystrophy, hyperlipidaemia and insulin resistance in patients receiving HIV protease inhibitors. AIDS 1998; 12:F51-F58.

(226) Dronda F, Moreno S, Perez-Elias MJ, Antela A, Moreno A. Vascular disease in HIV-infected patients: a comparative study of two different theraputic periods (1994-1997 versus 1998-2000). AIDS 2002; 16(14):1971-1973.

(227) Coplan PM, Nikas A, Japour A, Cormier K, Maradit-Kremers H, Lewis R et al. Incidence of myocardial infarction in randomised clinical trials of protease inhibitor-based antiretroviral therapy: an analysis of four different protease inhibitors. AIDS Res Hum Retroviruses 2003; 19(6):449-455.

(228) Bozzette SA, Ake CF, Tam HK, Chang SW, Louis TA. Cardiovascular and cerebrovascular events in patients treated for human immunodeficiency virus infection. The New England Journal of Medicine 148[8], 702-710. 20-2-2003. Ref Type: Abstract

(229) Koppel K, Bratt G, Schulman S, Bylund H, Sandstrom E. Hypofibrinolytic state in HIV-1-infected patients treated with protease inhibitor-containing highly active antiretroviral therapy. Journal of Acquired Immune Deficiency Syndrome 2001; 29:441-449.

(230) Lichtenstein KA, Delaney KM, Armon C, Ward DJ, Moorman AC, Wood KC et al. Incidence of and risk factors for lipoatrophy (abnormal fat loss) in ambulatory HIV-1 infected patients. J Acquir Immune Defic Syndr Hum Retrovirol 2003; 32:48-56.

(231) Safrin S, Grunfeld C. Fat distribution and metabolic changes in patients with HIV infection. AIDS 1999; 13(18).

(232) Rhew DC, Bernal M, Aguilar D, Iloeje U, Goetz MB. Association between Protease Inhibitor Use and Increased Cardiovascular Risk in Patients Infected with Human Immunodeficiency Virus: A Systematic Review. Clin Infect Dis 2003; 37(7):959-972.

(233) Graham NM. Metabolic disorders among HIV-infected patients treated with protease inhibitors: a review. J Acquir Immune Defic Syndr 2000; 25:S4-S11.

(234) Wanke CA. Epidemiological and clinical aspects of the metabolic complications of HIV infection: The fat redistribution syndrome. AIDS 1999; 13(11):1287-1293.

(235) John M, Nolan D, Mallal S. Editorial Review: Anitretroviral therapy and the lipodystrophy syndrome. Antivir Ther 2001; 6:9-20.

(236) Drechsler H, Powderly WG. Switching effective antiretroviral therapy: a review. Clin Infect Dis 2002; 35(10):1219-1230.

(237) Murphy RL, Smith WJ. Switch studies: A review. HIV Med 2002; 3:146-155.

(238) Kamin DS, Grinspoon SK. Cardiovascular disease in HIV-positive patients. AIDS 2005; 19(7):641-652.

(239) Martinez E, Tuset M, Milinkovic A, Miro JM, Gatell JM. Management of dyslipidaemia in HIV-infected patients receiving antiretroviral therapy. Antivir Ther 2004; 9(5):649-663.

(240) Schambelan M, Benson CA, Carr A, Currier JS, Dube MP, Gerber JG et al. Management of Metabolic Complications Associated With Antiretroviral Therapy for HIV-1 Infection: Recommendations of an International AIDS Society–USA Panel. J Acquir Immune Defic Syndr 2002; 31(3):257-275.

(241) Grunfeld C, Pang M, Doerrler W, Shigenaga JK, Jensen P, Feingold KR. Lipids, lipoproteins, triglyceride clearence and cytokines in human immunodeficiency virus infection and the acquired immunodeficiency syndrome. J Clin Endocrinol Metab 1992; 74(5):74-1045.

(242) Behrens GM, Meyer-Olson D, Stoll M, Schmidt RE. Clinical impact of HIV-related lipodystrophy and metabolic abnormalities on cardiovascular disease. AIDS 2003; 17(Suppl 1):S149-S154.

(243) Grunfeld C, Kotler DP, Hamadeh R, Tierney A, Wang J, Pierson RN. Hypertriglyceridaemia in the acquired immunodeficiency syndrome. Am J Med 1989; 86(1):27-31.

(244) Rodriguez-Guardado A, Maradona JA, Carton JA, Asensi V. Triglyceride increase can predict lipodystrophy in HIV patients under highly active antiretroviral therapy. AIDS 2003; 16(10):1434-1436.

(245) Floris-Moore M, Howard AA, Lo Y, Arnsten JH, Santoro N, Schoenbaum EE. Increased serum lipids are associated with higher CD4 lymphocyte count in HIV-infected women. HIV Med 2006; 7(7):421-430.

(246) The Strategies for Management of Antiretroviral Therapy (SMART) Study Group. CD4+ count-guided interruption of antiretroviral treatment. N Eng J Med 2006; 355(22):2283-2296.

(247) Mary-Krause M, Cotte L, Partisani M, Simon A, Costagliola D. Impact of treatment with protease inhibitor (PI) in myocardial infarction (MI) occurrence in HIV-infected men. 8th Conference on Retroviruses and Opportunistic Infections 2001;Abstract 657.

(248) Braitstein P, Yip B, Heath KV, Levy AR, Montaner JSG, Humphries K et al. Interventional cardiovascular procedures among HIV-infected individuals on antiretroviral therapy 1995-2000. AIDS 2003; 17(14):2071-2075.

(249) Visnegarwala F, Darcourt J, Sajja P, Menon V, Ong O, Maldonado M et al. Changes in metabolic profile among antiretroviral-naive patients initiating protease inhibitor versus no-protease inhibitor containing HAART regimens. J Acquir Immune Defic Syndr 2003; 33(5):653-654.

(250) Heath KV, Chan KJ, Singer J, O'Shaughnessy MV, Montaner JSG, Hogg RS. Incidence of morphological and lipid abnormalities: gender and treatment differentials after initiation of first antiretroviral therapy. Int J Epidemiol 2002; 31(5):1016-1020.

(251) Martinez E, Arnaiz JA, Podzamczer D, Dalmau D, Ribera E, Domingo P et al. Substitution of nevirapine, efavirinz, or abacavir for Protease inhibitors in patients with human immunodeficiency virus infection. N Eng J Med 2003; 349(11):1036-1046.

(252) Ruiz L, Negredo E, Domingo P, Parades R, Francia E, Balague M et al. Antiretroviral treatment simplification with nevirapine in protease inhibitor-experienced patients with HIV-associated lipodystrophy: 1- year prospective follow-up of a multicenter, randomised, controlled study. J Acquir Immune Defic Syndr 2001; 27(3):229-236.

(253) Thiebaut R, Dabis F, Malvy D, Jacqmin-Gadda H, Mercie P, Daucourt V. Serum triglycerides, HIV infection and highly active antiretroviral therapy, Acquitaine Cohort, France, 1996 to 1998. J Acquir Immune Defic Syndr 2000; 23:261-265.

(254) Smith CJ, Levy I, Sabin CA, Kaya E, Johnson MA, Lipman MCI. Cardiovascular disease risk factors and antiretroviral therapy in an HIV-positive UK population. HIV Med 2003;submitted.

(255) Hodgson C, Hodgson L, Murad S, Herman J, Taylor C, Macallan D et al. Prevalence of cardiovascular disease (CVD) risk factors in an ethnically diverse HIV-positive population. 8th Annual conference of the British HIV association (BHIVA) , Abstract P2. 2002.
Ref Type: Abstract

(256) Khot UN, Khot MB, Bajzer CT, Sapp SK, Ohman EM, Ellis SG et al. Prevalence of conventional risk factors in patients with coronary heart disease. JAMA 2003; 290(898):904.

(257) Saves M, Chene G, Ducimetiere P, Leport C, Le Moal G, Amouyel P et al. Risk factors for coronary heart disease in patients treated for human

immunodeficiency virus infection compared with the general population. Clin Infect Dis 2003; 37:292-298.

(258) Eberly LE, Stamler J, Neaton JD. Relation of triglyceride levels, fasting and nonfasting, to fatal and nonfatal coronary heart disease. Arch Intern Med 2003; 163:1077-1083.

(259) Matthews GV, Moyle GJ, Mandalia S, Bower M, Nelson M, Gazzard BG. Absence of association between individual thymidine analogues or nonnucleoside analogues and lipid abnormalities in HIV-1 infected persons on initial therapy. J Acquir Immune Defic Syndr 2000; 24:310-315.

(260) Carter VM, Hoy JF, Bailey M, Colman PG, Myulasi I, Mijch AM. The prevalence of lipodystrophy in an ambulant HIV-infected population: it all depends on the definition. HIV Med 2001; 2(3):174-180.

(261) Law M, Puls R, Cheng AK, Cooper DA, Carr A. Evaluation of the HIV lipodystrophy case definition in a placebo-controlled, 144-week study in antiretroviral-naive adults. Antivir Ther 2006; 11(2):179-186.

(262) Yang Y, Zhu WDJ, Paton NI. Comparison of dual-energy X-ray absorptiometry machines for measuring fat distribution changes of HIV-associated lipodystrophy. Antivir Ther 2004; 9(5):771-778.

(263) Carr A, Emery S, Law M, Puls R, Lundgren JD, Powderly WG et al. An objective case definition of lipodystrophy in HIV-infected adults: a case-control study. Lancet 2003; 361(9359):726-735.

(264) Ioannidis JPA, Trikalinos TA, Law M, Carr A, for the HIV Lipodystrophy Case Definition Study Group. HIV lipodystrophy case definition using artificial neural network modelling. Antivir Ther 2003; 8(5):435-441.

(265) Iloeje UH, Yuan Y, L'Italien G, Mauskopf J, Holmberg SD, Moorman AC et al. Protease inhibitor exposure and increased risk of cardiovascular disease in HIV-infected patients. HIV Med 2005; 6(1):37-44.

(266) The Data Collection on Adverse Events of Anti-HIV Drugs (DAD) Study Group. Class of Antiretroviral Drugs and the Risk of Myocardial Infarction. N Eng J Med 2007; 356:1723-1735.

(267) Domingo P, Labarga P, Palacios R, Guerro MF, Terron JA, Elias MJ et al. Improvement of dyslipidemia in patients switching from stavudine to tenofovir: preliminary results. AIDS 2004; 18(10):1475-1478.

(268) Gathe JC, Ive P, Wood R, Schurmann D, Bellos NC, DeJesus E et al. SOLO: 48-week efficacy and safety comparison of once-daily fosamprenavir /ritonavir versus twice-daily nelfinavir in naive HIV-1-infected patients. AIDS 2004; 18(11):1529-1537.

(269) Llibre JM, Domingo P, Palacios R, Santos J, Perez-Elias MJ, Sanchez-de la Rosa R et al. Sustained improvement of dyslipidaemia in HAART-treated patients replacing stavudine with tenofovir. AIDS 2006; 20(10):1407-1414.

(270) Calza L, Manfredi R, Colangeli V, Tampellini L, Sebastiani T, Pocaterra D et al. Substitution of nevirapine or efavirenz for protease inhibitor versus lipid-lowering therapy for the management of dyslipidaemia. AIDS 2005; 19(10):1051-1058.

(271) Young J, Weber R, Rickenbach M, Furrer H, Bernasconi E, Hirschel B et al. Lipid profiles for antiretroviral-naive patients starting PI- and NNRTI-based therapy in the Swiss HIV Cohort Study. Antivir Ther 2005; 10(5):585-591.

(272) Jones R, Sawleshwarkar S, Michailidis C, Jackson A, Mandalia S, Stebbing J et al. Impact of antiretroviral choice on hypercholesterolaemia events: the role of the nucleoside reverse transcriptase inhibitor backbone. HIV Med 2005; 6(6):396-402.

(273) van Leth F, Hall DB, Lange JMA, Reiss P. Plasma lipid concentrations after 1.5 years of exposure to nevirapine or efavirenz together with stavudine and lamivudine. HIV Med 2006; 7(6):347-350.

(274) Richter A, Pladevall M, Manjunath R, Lafata JE, Xi H, Simpkins J et al. Patient characteristics and costs associated with dyslipidaemia and related conditions in HIV-infected patients: a retrospective cohort study. HIV Med 2005; 6(2):79-90.

(275) Shikuma C, Yang Y, Glesby M, Meyer W, Tashima KT, Ribaudo HJ et al. Metabolic Effects of Protease Inhibitor-Sparing Antiretroviral Regimens Given as Initial Treatment of HIV-1 Infection (AIDS Clinical Trials Group Study A5095). J Acquir Immune Defic Syndr 2007; 44(5):540-550.

(276) Fontas E, van Leth F, Sabin CA, Friis-Moller N, Rickenbach M, d'Arminio Monforte A et al. Lipid profiles in HIV-infected patients receiving combination antiretroviral therapy: are different antiretroviral drugs associated with different lipid profiles? J Infect Dis 2004; 189:1056-1074.

(277) Young J, Rickenbach M, Weber R, Furrer H, Bernasconi E, Hirschel B et al. Body fat changes among antiretroviral-naive patients on PI- and NNRTI-based HAART in the Swiss HIV Cohort Study. Antivir Ther 2005; 10(1):73-81.

(278) Seminari E, Tinelli C, Minoli L, Sacchi P, Filice G, Zocchetti C et al. Evaluation of the risk factors associated with lipodystrophy development in a cohort of HIV-positive patients. Antivir Ther 2002; 7(3):175-180.

(279) Karmon SL, Moore RD, Dobs AS, Keruly J, Barnett S, Cofrancesco J. Body shape and composition in HIV-infected women: an urban cohort. HIV Med 2005; 6(4):245-252.

(280) Bernasconi E, Boubaker K, Junghans C, Flepp M, Furrer H, Haensel A et al. Abnormalities of Body Fat Distribution in HIV-Infected Persons Treated With Antiretroviral Drugs: The Swiss HIV Cohort Study. J Acquir Immune Defic Syndr 2002; 31(1):50-55.

(281) Cameron DW, da Silva B, Arribas J, Pulido F, Katner H, Wikstrom K et al. Significant Sparing of Peripheral Lipoatrophy by HIV Treatment with LPV/r + ZDV/3TC Induction followed by LPV/r Monotherapy Compared with EFV + ZDV/3TC. 14th Conference on Retrovirus and Opportunistic Infections 2007; 44.

(282) Tien PC, Cole SR, Williams CM, Li R, Justman JE, Cohen MH et al. Incidence of lipoatrophy and lipohypertrophy in the women's interagency HIV study. J Acquir Immune Defic Syndr 2003; 34(5):461-466.

(283) Bonfanti P, Gulisano C, Ricci E, Timillero L, Valsecchi L, Carradori S et al. Risk factors for lipodystrophy in the CISAI cohort. Biomedecine & Pharmacotherapy 2003; 57(9):422-427.

(284) Rodriguez-Guardado A, Maradona JA, Carton JA, Asensi V. Triglyceride increase can predict lipodystrophy in HIV patients under highly active antiretroviral therapy. AIDS 2002; 16(10):1434-1436.

(285) Dube MP, Parker RA, Tebas P, Grinspoon S, Zackin R, Robbins GK et al. Glucose metabolism, lipid, and body fat changes in antiretroviral-naive subjects randomized to nelfinavir or efavirenz plus dual nucleosides. AIDS 2005; 19(16):1807-1818.

(286) Moyle GJ, Sabin CA, Cartledge J, Johnson M, Wilkins E, Churchill D et al. A randomized comparative trial of tenofovir DF or abacavir as replacement for a thymidine analogue in persons with lipoatrophy. AIDS 2006; 20(16):2043-2050.

(287) Domingo P, Sambeat MA, Perez A, Ordonez J, Rodriguez J, Vazquez G. Fat distribution and metabolic abnormalities in HIV-infected patients on first combination antiretroviral therapy including stavudine. Antivir Ther 2003; 8(3):223-231.

(288) Podzamczer D, Ferrer E, Sanchez P, Gatell JM, Crespo M, Fisac C et al. Less lipoatrophy and better lipid profile with abacavir as compared to stavudine: 96-week results of a randomized study. J Acquir Immune Defic Syndr 2007; 44(2):139-147.

(289) Haubrich RH, Riddler S, DiRienzo G, Komarow L, Powderly W, Garren K et al. Metabolic Outcomes of ACTG 5142: A Prospective, Randomized, Phase III Trial of NRTI-, PI-, and NNRTI-sparing Regimens for Initial Treatment of HIV-1 Infection. 14th Conference on Retrovirus and Opportunistic Infections 2007; 38.

(290) Stone SF, Lee S, Keane NM, Price P, French MA. Association of increased hepatitis C virus (HCV)-specific IgG and soluble CD26 dipeptidyl peptidase IV enzyme activity with hepatotoxicity after highly active antiretroviral therapy in human immunodeficiency virus-HCV-coinfected patients. J Infect Dis 2002; **186**(**10**):1498-1502.

(291) De Simone JA, Pomerantz RJ, Babinchak TJ. Inflammatory reactions in HIV-1-infected persons after initiation of highly active antiretroviral therapy. Ann Intern Med 2000; 133(6):447-454.

(292) Pomerantz RJ. The expert speaks: Immune reconstitution in AIDS. AIDS Patient Care STDS 2003; 17(3):99-101.

(293) Martin-Carbonero L, Soriano V. Interplay between hepatitis C, liver steatosis and antiretroviral therapy in HIV-infected patients. AIDS 2005; 19(6):621-623.

(294) Puoti M, Torti C, Ripamonti D, Castelli F, Zaltron S, Zanini B et al. Severe hepatotoxicity during combination antiretroviral therapy treatment:incidence, liver histology, and outcome. J Acquir Immune Defic Syndr 2003; 32(3):259-267.

(295) Saves M, Vandentorren S, Daucourt V, Marimoutou C, Dupon M, Couzigou P et al. Severe hepatic cytolysis: incidence and risk factors in patients treated by antiretroviral combinations Aquitaine Cohort, France, 1996-1998. AIDS 1999; 13(17):115-151.

(296) Gisolf EH, Dreezen C, Danner SA, Weel J, Weverling GJ. Risk factors for hepatotoxicity in HIV-1-infected patients receiving Ritonavir and Saquinavir with or without Stavudine. Clin Infect Dis 2000; 31:1234-1239.

(297) d'Arminio Monforte A, Bugarini R, Pezzotti P, de Luca A, Antinori A, Mussini C et al. Low frequency of severe hepatotoxicity and association with HCV coinfection in HIV-positive patients treated with HAART. J Acquir Immune Defic Syndr Hum Retrovirol 2001; 28(2):114-123.

(298) Aceti A, Pasquazzi C, Zechini B, De Bac C. Hepatotoxicity development during antiretroviral therapy containing protease inhibitors in patients with HIV. J Acquir Immune Defic Syndr Hum Retrovirol 2002; 29:41-48.

(299) Nunez M, Lana R, Mendoza JL, Martin-Carbonero L, Soriano V. Risk factors for severe hepatic injury after introduction of highly active antiretroviral therapy. J Acquir Immune Defic Syndr Hum Retrovirol 2001; 27(5):426-431.

(300) Bonfanti P, Landonio S, Ricci E, Martinelli C, Fortuna P, Faggion I et al. Risk factors for hepatotoxicity in patients treated with highly active antiretroviral therapy. J Acquir Immune Defic Syndr Hum Retrovirol 2001; 27(3):316-318.

(301) Sulkowski MS, Thomas DL, Chaisson RE, Moore RD. Hepatotoxicity associated with antiretroviral therapy in adults infected with Human Immunodeficiency Virus and the role of Hepatitis C or B virus infection. JAMA 2000; 283(1):74-80.

(302) den Brinker M, Wit FWNM, Wertheim-van Dillen PME, Jurriaans S, Weel J, van Leeuwen DM et al. Hepatitus B and C virus co-infection and the risk for hepatotoxicity of highly active antiretroviral therapy in HIV-1 infection. AIDS 2000; 14(2895):2902.

(303) Martinez E, Blanco JL, Arnaiz JA, Perez-Cuevas JB, Mocroft A, Cruceta A et al. Hepatotoxicity in HIV-1 infected patients receiving nevirapine-containing antiretroviral therapy. AIDS 2001; 15:1261-1268.

(304) Saves M, Raffi F, Clevenbergh P, Marchou B, Walder-Combernoux A, Morlat P et al. Hepatitis B or Hepatitis C virus infection is a risk factor for severe hepatic cytolysis after initiation of a protease inhibitor-containing antiretroviral regimen in human immunodeficiency virus-infected patients. Antimicrob Agents Chemother 2000; 44(12):3451-3455.

(305) Wit FWNM, Weverling GJ, Weel J, Jurriaans S, Lange JMA. Incidence of and risk factors for severe hepatotoxicity associated with antiretroviral combination therapy. J Infect Dis 2002; 186:22-31.

(306) Sulkowski MS, Thomas DL, Mehta SH, Chaisson RE, Moore RD. Hepatotoxicity associated with Nevirapine or Efavirenz-containing antiretroviral therapy: role of Hepatitis C and B infections. Hepatology 2002; 35:182-189.

(307) Palmon R, Koo BCA, Shoultz DA, Dieterich DT. Lack of hepatotoxicity associated with nonnucleoside reverse transcriptase inhibitors. J Acquir Immune Defic Syndr Hum Retrovirol 2002; 29:340-345.

(308) Thio CL, Seaberg EC, Skolasky R, Phair JP, Visscher B, Munoz A et al. HIV-1, hepatitis B virus, and risk of liver-related mortality in the Multicenter Cohort Study (MACS). Lancet 2002; 360(1921):1926.

(309) AIDS Clinical Trials Group. Table of grading and severity of adult adverse experiences. Rockville: Division of AIDS 1996; National Institute of Allergy and Infectious Diseases.

(310) Sulkowski MS, Moore RD, Mehta SH, Chaisson RE, Thomas DL. Hepatitus C and progression of HIV disease. JAMA 2002; 288(2):199-206.

(311) Centers for Disease Control and Prevention. Recommendations for prevention and control of hepatitis C virus (HCV) infection and HCV-related chronic disease. MMWR Morb Mortal Wkly Rept 1998; 47(RR19):1-39.

(312) Pirisi M, Toniutto P, Fabris C, Falleti E, Tisminetzky SG, Baralle F et al. Factors associated with serum HCV RNA positivity in anti-HCV antibody positive intravenous drug users. J Clin Epidemiol 1998; 51(5):423-427.

(313) Barrett S, Kieran N, Ryan E, O'Keane JC, Crowe J. Intrahepatic hepatitis C viral RNA status of serum polymerase chain reaction-negative individuals with histological changes on liver biopsy. Hepatology 2001; 33(6):1496-1502.

(314) Martin-Carbonero L, Nunez M, Rios P, Perez-Olmeda M, Gonzalez-Lahoz J, Soriano V. Liver injury after beginning antiretroviral therapy in HIV/hepatitis C virus co-infected patients is not related to immune reconstitution. AIDS 2002; 16(10):1423-1425.

(315) Macias J, Castellano V, Merchante N, Palacios R, Mira JM, Saez C et al. Effect of antiretroviral drugs on liver fibrosis in HIV-infected patients with chronic hepatitis C: harmful impact of nevirapine. AIDS 2004; 18(5):767-774.

(316) Sulkowski MS, Mehta SH, Chaisson RE, Thomas DL, Moore RD. Hepatotoxicity associated with protease inhibitor-based antiretroviral regimens with or without concurrent ritonavir. AIDS 2004; 18(17):2277-2284.

(317) Bonfanti P, Ricci E, Penco G, Orofino G, Bini T, Sfara C et al. Low incidence of hepatotoxicity in a cohort of HIV patients treated with lopinavir/ritonavir. AIDS 2005; 19(13):1433-1434.

(318) Mocroft A, Soriano V, Rockstroh J, Reiss P, Kirk O, De Wit S et al. Is there evidence for an increase in the death rate from liver-related disease in patients with HIV? AIDS 2005; 19(18):2117-2125.

(319) Manfredi R, Calza L. Nevirapine versus efavirenz in 742 patients: no link of liver toxicity with female sex, and a baseline CD4 cell count greater than 250 cells/µl. AIDS 2006; 20(17):2233-2236.

(320) Meraviglia P, Schiavini M, Castagna A, Vigano P, Bini T, Landonio S et al. Lopinavir/ritonavir treatment in HIV antiretroviral-experienced patients: evaluation of risk factors for liver enzyme elevation. HIV Med 2004; 4(4):334-343.

(321) Maida I, Nunez M, Rios MJ, Martin-Carbonero L, Sotgiu G, Toro C et al. Severe Liver Disease Associated With Prolonged Exposure to Antiretroviral Drugs. J Acquir Immune Defic Syndr 2006; 42(2):177-182.

(322) Sanne I, Mommeja-Marin H, Hinkle J, Bartlett JA, Lederman MM, Maartens G et al. Severe Hepatotoxicity Associated with Nevirapine Use in HIV-Infected Subjects. J Infect Dis 2005; 191(6):825-829.

(323) Barreiro P, Maida I, Martin-Carbonero L, Blanco F, Labarga P, Castellares C et al. Liver Disease of Unknown Etiology in HIV-infected Patients: Prevalence, Clinical Presentation, and Putative Causes. 14th Conference on Retrovirus and Opportunistic Infections 2007; 821.

(324) Maggiolo F, Arici C, Airoldi M, Ripamonti D, Quinzan G, Gregis G et al. Reasons for discontinuation of nevirapine-containing HAART: results from an

unselected population of a large clinical cohort. J Antimicrob Chemother 2007; 59(3):569-572.

(325) Scheuer PJ. Classification of chronic viral hepatitis: a need for reassessment. J Hepatol 1991; 13:372-374.

(326) Laradi A, Mallet A, Beaufils H, Allouache M, Martinez F. HIV-associated nephropathy: outcome and prognosis factors. Groupe d'Etudes Nephrologiques d'Ile de France. J Am Soc Nephrol 1998; 9(12):2327-2335.

(327) Betjes MGH, Weening J, Krediet RT. Diagnosis and treatment of HIV-associated nephropathy. Neth J Med 2001; 59:111-117.

(328) Kimmel PL. The nephropathies of HIV infection:pathogenesis and treatment. Curr Opin Nephrol Hypertens 2000; 9:117-122.

(329) Krawczyk CS, Holmberg SD, Moorman AC, Gardner LI, McGwin G, for the HIV Outpatient Study Group. Factors associated with chronic renal failure in HIV-infected ambulatory patients. AIDS 2004; 18(16):2171-2178.

(330) Izzedine H, Deray G. The nephrologist in the HAART era. AIDS 2007; 21(4):409-421.

(331) Grabe DW, Eisele G, Miller C, Singh J, Stein D. Indinavir-induced nephropathy. Clin Nephrol 1999; 51(3):181-183.

(332) Guery B, Tubiana R, Martinez F, Jacquiaud C, Bochet M, Katlama C. Renal tolerance of indinavir in HIV-positive patients. Nephron 1999; 82(1):72.

(333) Dieleman JP, van Rossum AMC, Stricker BCH, Sturkenboom MCJM, de Groot R, Telgt D et al. Persistent leukocyturia and loss of renal function in a prospectively monitored cohort of HIV-infected patients treated with indinavir. J Acquir Immune Defic Syndr 2003; 32:135-142.

(334) Deray G, Bochet M, Katlama C, Bricaire F. Nephrotoxicity of ritonavir. Presse Med 1998; 27(35):1801-1803.

(335) Karras A, Lafaurie M, Furco A, Bourgarit A, Droz D, Sereni D et al. Tenofovir-related nephrotoxicity in human immunodeficiency virus-infedcted patients: three cases of renal failure, Fanconi syndrome, and nephrogenic diabetes insipidus. Clin Infect Dis 2003; 36:1070-1073.

(336) Gallant JE. Similar 96-week renal safety profile of tenofovir DF versus stavudine (d4T) when used in combination with lamivudine (3TC) and efavirinz in antiretroviral naive patients. 43rd ICAAC, Chicago, 14-17th September 2003;abstract H-840.

(337) Cirillo M, Anastasio P, De Santo NG. Relationship of gender, age, and body mass index to errors in predicted kidney function. Nephrol Dial Transplant 2005; 20(9):1791-1798.

(338) Izzedine H, Isnard-Bagnis C, Hulot JS, Vittecoq D, Cheng A, Jais CK et al. Renal safety of tenofovir in HIV treatment-experienced patients. AIDS 2004; 18(7):1074-1076.

(339) Julg BD, Bogner JR, Crispin A, Goebel FD. Progression of renal impairment under therapy with tenofovir. AIDS 2005; 19(12):1332-1333.

(340) Wyatt CM, Arons RR, Klotman PE, Klotman ME. Acute renal failure in hospitalized patients with HIV: risk factors and impact on in-hospital mortality. AIDS 2006; 20(4):561-565.

(341) El Sahly HM, Teeter L, Zerai T, Andrade RA, Munoz C, Nnabuife C et al. Serum creatinine changes in HIV-seropositive patients receiving tenofovir. AIDS 2006; 20(5):786-787.

(342) Jaroszewicz J, Wiercinska-Drapalo A, Lapinski TW, Prokopowicz D, Rogalska M, Parfieniuk A. Does HAART improve renal function? An association between serum cystatin C concentration, HIV viral load and HAART duration. Antivir Ther 2006; 11(5):641-645.

(343) Gerard L, Chazallon C, Taburet AM, Girard PM, Aboulker JP, Piketty C. Renal function in antiretroviral-experienced patients treated with tenofovir disoproxil fumarate associated with atazanavir/ritonavir. Antivir Ther 2007; 12(1):31-39.

(344) Antoniou T, Raboud JM, Chirhin S, Yoong D, Govan V, Gough K et al. Incidence of and risk factors for tenofovir-induced nephrotoxicity: a retrospective cohort study. HIV Med 2005; 6(4):284-290.

(345) Jones R, Stebbing J, Nelson M, Moyle G, Bower M, Mandalia S et al. Renal Dysfunction With Tenofovir Disoproxil Fumarate-Containing Highly Active Antiretroviral Therapy Regimens Is Not Observed More Frequently: A Cohort and Case-Control Study. J Acquir Immune Defic Syndr 2004; 37(4):1489-1495.

(346) Moreno S, Domingo P, Palacios R, Santos J, Falco V, Murillas J et al. Renal Safety of Tenofovir Disoproxil Fumarate in HIV-1 Treatment-experienced Patients with Adverse Events Related to Prior NRTI Use: Data from a Prospective, Observational, Multicenter Study. J Acquir Immune Defic Syndr 2006; 42(3):385-387.

(347) Buchacz K, Young B, Baker RK, Moorman A, Chmiel JS, Wood KC et al. Renal Function in Patients Receiving Tenofovir With Ritonavir/Lopinavir or Ritonavir/Atazanavir in the HIV Outpatient Study (HOPS) Cohort. J Acquir Immune Defic Syndr 2006; 43(5):626-628.

(348) Overton E, Nurutdinova D, Freeman J, Finn D, Seyfried W, Mondy K. Factors Associated with Renal Dysfunction within an Urban Cohort in the Era of HAART. 14th Conference on Retrovirus and Opportunistic Infections 2007; P826.

(349) Mocroft A, Kirk O, Gatell J, Reiss P, Gargalianos-Kakolyris P, Zilmer K et al. Detection of Chronic Renal Failure among HIV Patients within the EuroSIDA Study. 14th Conference on Retrovirus and Opportunistic Infections 2007; 828.

(350) Goicoechea M, Liu S, Best B, Sun S, Jain S, Kemper C et al. Increased Renal Impairment in Patients Receiving TDF+PI vs TDF+NNRTI. 14th Conference on Retrovirus and Opportunistic Infections 2007; 835.

(351) Garcia-Diaz A, Clewley GS, Booth CL, Labett W, McAllister N, Geretti AM. Comparative evaluation of the performance of the Abbott real-time human immunodeficiency virus type 1 (HIV-1) assay for measurement of HIV-1 plasma viral load following automated specimen preparation. J Clin Microbiol 2006; 44(5):1788-1791.

(352) Gibb DM, Duong T, Tookey PA, Sharland M, Tudor-Williams G, Novelli V et al. Decline in mortality, AIDS, and hospital admissions in perinatally HIV-1 infected

children in the United Kingdom and Ireland. Br Med J 2003; 327(7422):1019-1024.

(353) Lampe FC, Smith CJ, Madge S, Kinloch-de-Loes S, Tyrer M, Sabin CA et al. Success of HIV clinical care according to demographic group among sexually-infected patients in a routine clinic population, 1999 to 2004. Arch Intern Med 2007; In press.

(354) BHIVA writing committee on behalf of the British HIV association (BHIVA). BHIVA guidelines for the treatment of HIV-infected adults with antiretroviral therapy. HIV Med 2001; 1:76-101.

(355) Pozniak A, Gazzard B, Anderson J, Babiker A, Churchill D, Collins S et al. British HIV association (BHIVA) guidelines for the treatment of HIV-infected adults with antiretroviral therapy. HIV Med 2003; 4(Supp 1):1-41.

(356) The Antiretroviral Therapy (ART) Cohort Collaboration. Prognostic importance of initial response in HIV-1 infected patients starting potent antiretroviral therapy: analysis of prospective studies. Lancet 2003; 362:679-686.

(357) Bak Dragsted U, Gerstoft J, Pederson C, Peters B, Duran A, Obel N et al. Randomized trial to evaluate indinavir-ritonavir versus saquinavir/ritonavir in human immunodeficiency virus type 1-infected patients: the MaxCMin1 trial. J Infect Dis 2003; 186:635-642.

(358) Moyle GJ, Back D. Principles and practice of HIV-protease inhibitor pharmacoenhancement. HIV Med 2001; 2:105-13.

(359) http://www phls co uk/infections/default htm [2007

(360) Rodriguez-Novoa S, Barreiro P, Jimenez-Nacher I, Soriano V. Overview of the pharmacogenetics of HIV therapy. Pharmacogenomics J 2006; 6(4):234-245.

(361) Easterbrook PJ, Phillips AN, Hill T, Matthias R, Sabin C, Fisher M et al. Patterns and predictors of antiretroviral drug use at treatment initiation in the UK. XV International AIDS conference, Bangkok, 11-16 July 2004 2004; TuPeB4543.

(362) van Leth F, Phanuphak P, Stroes E, Gazzard B, Cahn P, Raffi F et al. Nevirapine and Efavirenz Elicit Different Changes in Lipid Profiles in Antiretroviral-Therapy-Naive Patients Infected with HIV-1. PLoS 2004; e19.

(363) Jousilahti P, Vartiainen E, Tuomilehto J, Puska P. Clinical Investigation and ReportsSex, Age, Cardiovascular Risk Factors, and Coronary Heart Disease A Prospective Follow-Up Study of 14 786 Middle-Aged Men and Women in Finland. Circulation 1999; 99:1165-1172.

(364) Korthuis PT, Asch SM, Anaya HD, Morganstern H, Bidwell Goetz M, Yano EM et al. Lipid screening in HIV-infected veterans. J Acquir Immune Defic Syndr 2004; 35(3):253-260.

(365) Clinical Trials: A Practical Guide to Design, Analysis, and Reporting. 1 ed. REMEDICA; 2005.

(366) Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc 1996; 91(434):473-489.

(367) Carrieri MP, Leport C, Protopopescu C, Cassuto JP, Bouvet E, Peyramond D et al. Factors associated with nonadherence to highly active antiretroviral therapy. A 5-year follow-up analysis with correction for the bias induced by missing data in the treatment maintenance phase. J Acquir Immune Defic Syndr 2006; 41(4):477-485.

(368) Touloumi G, Pocock SJ, Babiker AG, Darbyshire JH. Impact of missing data due to selective dropouts in cohort studies and clinical trials. Epidemiology 2002; 13:347-355.

(369) Gutierrez F, Navarro A, Padillia S, Anton R, Masia M, Borras J et al. Prediction of neuropsychiatric adverse events associated with long-term efavirenz therapy, using plasma drug level monitoring. Clin Infect Dis 2005; **41(11)**:1648-1653.

(370) Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. JASA 1999; 94:496-509.

(371) Levin J. Switching to lopinavir/ritonavir (LPV/r) Tablets once daily (TAB-QD) from soft-gel capsule dosed BID/QD (SGC-BID/SGC-QD) led to significant improvements in tolerability, diarrhea, antidiarrheal medication use, and satisfaction. 8th International Workshop on Adverse Drug Reactions and Lipodystrophy in HIVSan Francisco, Sept 24-26, 2006 2006.

(372) Conroy RM, Pyorala K, Fitzgerald AP, Sans S, Menotti A, De Backer G et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. European Heart Journal 2003; 24(11):987-1003.

(373) http://www heart nhs uk/scripts/default asp?site_id=23&id=28154 [2007

(374) Iribarren C, Sharp D, Burchfiel CM, Sun P, Dwyer JH. Association of Serum Total Cholesterol with Coronary Disease and All-Cause Mortality: Multivariate Correction for Bias Due to Measurement Error. Am J Epidemiol 1996; 143(5):463-471.

(375) Gallant JE. Initial therapy of HIV infection. Journal of clinical virology 2002; 25:317-333.

(376) Lichtenstein KA, Armon C, Baron A, Moorman AC, Wood KC, Holmberg SD et al. Modification of the Incidence of Drug-Associated Symmetrical Peripheral Neuropathy by Host and Disease Factors in the HIV Outpatient Study Cohort. Clin Infect Dis 2005; 40:148-157.

(377) Thiebaut R, Chene G, Jacqmin-Gadda H, Morlat P, Mercie P, Dupon M et al. Time-updated CD4+ T lymphocyte count and HIV RNA as major markers of disease progression in naive HIV-1-infected patients treated with a highly active antiretroviral therapy: the Aquitaine cohort, 1996-2001. J Acquir Immune Defic Syndr 2003; 33(3):380-386.

(378) Wit FW, Wood R, Horban A, Beniowski M, Schmidt RE, Gray G et al. Prednisolone does not prevent hypersensitivity reactions in antiretroviral drug regimens containing abacavir with or without nevirapine. AIDS 2001; 15(18):2423-2429.

(379) Bonnet F, Bonarek M, Morlat P, Mercie P, Dupon M, Gemain MC et al. Risk factors for lactic acidosis in HIV-infected patients treated with nucleoside reverse transcrpitase inhibitors: a case control study. Clin Infect Dis 2003; 36:1324-1328.

(380) Bonnet F, Saves M, Droz C, Peuchant E, Chene G, Beylot J et al. Increase of atherogenic plasma profile in HIV-infected patients treated with protease inhibitor-containing regimens. J Acquir Immune Defic Syndr 2000; 25(2):199-200.

(381) Kirk O, Gerstoft J, Pederson C, Nielson H, Obel N, Katzenstein TL et al. Low body weight and type of protease inhibitor predict discontinuation and treatment-limiting adverse drug reactions among HIV-infected patients starting a protease inhibitor regimen: consistent results from a randomised trial and an observational cohort. HIV Med 2001; 2:43-51.

(382) Pintilie M. Analysing and interpreting competing risk data. Stat Med 2007; 26:1360-1367.

(383) Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. Stat Med 2007; 26:2389-2430.

(384) Pintilie M. Competing risks. A pratical perspective. 1st ed. Chichester, West Sussex, England: John Wiley and Sons Ltd; 2006.

(385) Qurishi N, Kreuzberg C, Luchters G, Effenberger W, Kupfer B, Sauerbruch T et al. Effect of antiretroviral therapy on liver-related mortality in patients with HIV and hepatitis C virus coinfection. Lancet 2003; 362:1708-1713.

(386) Dore G. Antiretroviral therapy-related hepatotoxicity: predictors and clinical management. J HIV Ther 2003; 8(4):96-100.

(387) Moyle G, Sawyer W, Law M, Amin J, Hill A. Changes in hematologic parameters and efficacy of thymidine analogue-based, highly active antiretroviral therapy: A meta-analysis of six prospective, randomized, comparative studies. Clin Ther 2004; 26(1):92-97.

(388) Lichtenstein KA, Ward DJ, Moorman AC, Delaney KM, Young B, Palella FJ et al. Clinical assessment of HIV-associated lipodystrophy in an ambulatory population. AIDS 2001; 15(11):1389-1398.

(389) Clayton D, Hills M. Statistical Models in Epidemiology. Oxford: Oxford University Press; 2001.

(390) Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika 1983; 70(1):41-55.

(391) de Luca A, Cozzi-Lepri A, Antinori A, Zaccarelli M, Bongiovanni M, Di Giambenedetto S et al. Lopinavir/ritonavir or efavirenz plus two nucleoside analogues as first-line antiretroviral therapy: a non-randomized comparison. Antivir Ther 2006; 11(5):609-618.

(392) Johnson M, Grinsztejn B, Rodriguez C, Coco J, DeJesus E, Lazzarin A et al. 96-week comparison of once-daily atazanavir/ritonavir and twice-daily lopinavir/ritonavir in patients with multiple virologic failures. AIDS 2006; 20(5):711-718.

(393) Montes ML, Pulido F, Barros C, Condes E, Rubio R, Cepeda C et al. Lipid disorders in antiretroviral-naive patients treated with lopinavir/ritonavir-based HAART: frequency, characterization and risk factors. J Antimicrob Chemother 2005; 55(5):800-804.

(394) Fumaz CR, Munoz-Moreno JA, Molto J, Negredo E, Ferrer MJ, Sirera G et al. Long-term neuropsychiatric disorders on efavirenz-based approaches: quality of life, psychologic issues, and adherence. J Acquir Immune Defic Syndr 2005; 38(5):560-565.

(395) Shelton BJ, Gilbert GH, Lu Z, Bradshaw P, Chavers LS, Howard G. Comparing longitudinal binary outcomes in an observational oral health study. Stat Med 2003; 22:2057-2070.

(396) Crown WH, Obenchain RL, Engelhart L, Lair T, Buesching DP, Croghan T. The application of sample selection models to outcomes research: the case of evaluating the effects of antidepressant therapy on resource utilization. Stat Med 1998; 17:1943-1958.

(397) Heckman JJ. Sample Selection bias as a specification error. Econometrica 1979; 47(1):153-162.

(398) Dolton PJ, Makepeace GH. Sample selection and male-femal earnings differentials in the graduate labour market. Oxford Economic Papers 1986; 38:317-341.

(399) Cheung YB. Adjustment for selection bias in cohort studies. An application of a probit model with selectivity to life course epidemiology. J Clin Epidemiol 2001; 54(12):1238-1243.

(400) Sales AE, Plomondon ME, Magid DJ, Spertus JA, Rumsfeld JS. Assessing response bias from missing quality of life data: The Heckman method. Health Qual Life Outcomes 2004; 2:49.

(401) Campbell CI, Alexander JA. Health services for women in outpatient substance abuse treatment. Health Serv Res 2005; 40(3):781-810.

(402) Sullivan PW, Ghushchyan V. Cardiovascular risk factor clusters and employment in the United States. Value Health 2007; 10(S1):S52-S58.

(403) De Santis M, Carducci B, De Santis L, Cavaliere AF, Straface G. Periconceptional Exposure to Efavirenz and Neural Tube Defects. Arch Intern Med 2002; 162(3):355.

(404) Clark RA, Squires KE. Gender-specific considerations in the antiretroviral management of HIV-infected women. Expert Rev Anti Infect Ther 2005; 3(2):213-227.

(405) Bushway S, Johnson BD, Slocum LA. Is the magic still there? The use of the Heckman two-step correction for selection bias in criminology. J Quant Criminol 2007; 23:151-178.

(406) Stolzenberg RM, Relles DA. Theory testing in a world of constrained research design. The significance of Heckman's censored sampling bias correction for nonexperimental research. Sociol Meth Res 1990; 18(4):395-415.

(407) Sartori AE. An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions. Political Analysis 2003; 11(2):111-138.

(408) Winship C, Mare RD. Models for sample selection bias. Annu Rev Sociol 1992; 18:327-350.

(409) Rosenbaum PR. Observational Studies. New York, USA: Springer-Verlag; 1995.

(410) Stoker TM. Consistent estimates of scaled coefficients. Econometrica 1986; 54:1461-1481.

(411) Klein RW, Spady RS. An efficient semiparametric estimator of the binary response model. Morristown, NJ: Bell Communication Res.; 1987.

(412) Cosslett SR. Distribution-free maximum likelihood estimator of the binary choice model. Econometrica 1983; 51:765-781.

(413) Sabin CA, Smith CJ, Lampe FC, Youle M, Johnson MA, Chavers S et al. The use of propensity scores and sample selection models to remove the effects of measured and unmeasured confounding: a practical example. 10th International Workshop on HIV Observational Databases, 23-26 March 2006, Madrid, Spain 2006.

(414) The UK Collaborative HIV Cohort Steering Committee. The creation of a large UK-based multicentre cohort of HIV-infected individuals: The UK Collaborative HIV Cohort (UK CHIC) study. HIV Med 2004; 5(2):115-124.

(415) Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silberhatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998; 97:1837-1847.

(416) Berger S, Kugler JD, Thomas JA, Friedberg DZ. Sudden cardiac death in children and adolescents: Introduction and overview. Pediatr Clin N Am 2004; 51(5):1201-1209.

(417) Burke V. Obesity in childhood and cardiovascular risk. Clin Exp Pharmacol Physiol 2006; 33(9):831-837.

(418) Li R, Chen W, Srinivasan SR, Bond MG, Tang R, Urbina EM. Childhood cardiovascular risk factors and carotid vascular changes in adulthood: the Bogalusa Heart Study. JAMA 2003; 290(17):2271-2276.

(419) Juonala M, Jarviasalo MJ, Maki-Torkko N, Kahonen M, Viikari JS, Raitakari OT. Risk factors identified in childhood and decreased carotid artery elasticity in adulthood: the Cardiovascular Risk in Young Finns Study. Circulation 2005; 112(10):1486-1493.

(420) Li S, Chen W, Srinivasan SR, Berenson GS. Childhood blood pressure as a predictor of arterial stiffness in young adults: the bogalusa heart study. Hypertension 2004; 43(3):541-546.

(421) Brotons C, Ribera A, Perich RM, Abrodos D, Magana P, Pablo S et al. Worldwide distribution of blood lipids and lipoproteins in childhood and adolescence: a review study. Atherosclerosis 1998; 139(1):1-9.

(422) Boulton TJ, Magarey AM, Cockington RA. Serum lipids and apolipoproteins from 1 to 15 years: changes with age and puberty, and relationships with diet, parental cholesterol and family history of ischaemic heart disease. Acta paediatrica 1995; 84(10):1113-1118.

(423) Amaya RA, Kozinetz CA, McMeans A, Schwarzwald H, Kline MW. Lipodystrophy syndrome in human immunodeficiency virus-infected children. Pediatr Infect Dis J 2002; 21(5):405-410.

(424) European Paediatric Lipodystrophy Group. Antiretroviral therapy, fat redistribution and hyperlipidaemia in HIV-infected children in Europe. AIDS 2004; 18(10):1443-1451.

(425) Jaquet D, Levine M, Ortega-Rodriguez E, Faye A, Polak M, Vilmer E et al. Clinical and metabolic presentation of the lipodystrophic syndrome in HIV-infected children. AIDS 2000; 14(14):2123-2128.

(426) Farley J, Gona P, Crain M, Cervia J, Oleske J, Seage G et al. Prevalence of Elevated Cholesterol and Associated Risk Factors Among Perinatally HIV-Infected Children (4-19 years old) in Pediatric AIDS Clinical Trials Group 219C. J Acquir Immune Defic Syndr 2005; 38(4):480-487.

(427) Arpadi SM, Cuff PA, Horlick M, Wang J, Kotler DP. Lipodystrophy in HIV-Infected Children Is Associated With High Viral Load and Low CD4+-Lymphocyte Count and CD4+-Lymphocyte. J Acquir Immune Defic Syndr 2001; 27(1):30-34.

(428) Rhoads MP, Smith CJ, Tudor-Williams G, Kyd P, Walters S, Sabin CA et al. Effects of highly active antiretroviral therapy on paediatric metabolite levels. 7 2006; 1(16):24.

(429) Lainka E, Oezbek S, Falck M, Ndagijimana J, Niehues T. Marked dyslipidemia in human immunodeficiency virus-infected children on protease inhibitor-containing antiretroviral therapy. Pediatrics 2002; 110(5):e56.

(430) Cheseaux JJ, Jotterand V, Aebi C, Gnehm H, Kind C, Nadal D et al. Hyperlipidemia in HIV-infected children treated with protease inhibitors: relevance for cardiovascular diseases. J Acquir Immune Defic Syndr 2002; 30(3):288-293.

(431) Bitnun A, Sochett E, Babyn P, Holowka S, Stephens D, Read S et al. Serum lipids, glucose homeostasis and abdominal adipose tissue distribution in protease inhibitor-treated and naive HIV-infected children. AIDS 2003; 17(9):1319-1327.

(432) Bockhorst JL, Ksseiry I, Toye M, Chipkin SR, Stechenberg BW, Fisher DJ et al. Evidence of human immunodeficiency virus-associated lipodystrophy syndrome in children treated with protease inhibitors. Pediatr Infect Dis J 2003; 22(5):463-465.

(433) Melvin AJ, Lennon S, Mohan KM, Purnell JQ. Metabolic abnormalities in HIV type 1-infected children treated and not treated with protease inhibitors. AIDS Res Hum Retroviruses 2001; 17(12):1117-1123.

(434) Funk MB, Linde R, Wintergerst U, Notheis G, Hoffmann F, Schuster T et al. Preliminary experiences with triple therapy including nelfinavir and two reverse transcriptase inhibitors in previously untreated HIV-infected children. AIDS 1999; 13(13):1653-1658.

(435) McComsey GA, Leonard E. Metabolic complications of HIV therapy in children. AIDS 2004; 18(13):1753-1768.

(436) Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling (with disc.). Applied Statistics 1994; 43:429-467.

(437) Statistics with Confidence. 2nd ed. 2005.

(438) Sterne JA, May M, Bucher HC, Ledergerber B, Furrer HC, Cavassini M et al. HAART and the heart: changes in coronary risk factors and implications for coronary risk in men starting antiretroviral therapy. J Intern Med 2007; 261(3):255-267.

(439) European Collaborative Study. Level and pattern of HIV-1-RNA viral load over age: differences between girls and boys? AIDS 2002; 16(1):97-104.

(440) Moore L, Lavoie A, LeSage N, Abdous B, Bergeron E, Liberman M et al. Statistical validation of the Revised Trauma Score. J Trauma 2006; 60(2):305-311.

(441) Wang Z, Hoy WE. Population distribution of high sensitivity C-reactive protein in Aboriginal Australians: a comparison with other populations. Clin Biochem 2006; 39(3):277-281.

(442) Shkedy Z, Aerts M, Molenberghs G, Beutels P, Van Damme P. Modelling age-dependent force of infection from prevalence data using fractional polynomials. Stat Med 2006; 25(9):1577-1591.

(443) Moore L, Lavoie A, Camden S, LeSage N, Sampalis JS, Bergeron E et al. Statistical validation of the Glasgow Coma Score. J Trauma 2006; 60(6):1238-1243.

(444) Royston P, Reitz M, Atzpodien J. An approach to estimating prognosis using fractional polynomials in metastatic renal carcinoma. Br J Cancer 2006; 94(12):1785-1788.

(445) Khoshdel AR, Thakkinstian T, Carney SL, Attia J. Estimation of an age-specific reference interval for pulse wave velocity: a meta-analysis. J Hypertens 2006; 24(7):1231-1237.

(446) Polesel J, Dal Maso V, Bagnardi V, Zucchetto A, Zambon A, Levi F et al. Estimating dose-response relationship between ethanol and risk of cancer using regression spline models. Int J Cancer 2005; 114(5):836-841.

(447) Blair EM, Liu Y, de Klerk NH, Lawrence DM. Optimal fetal growth for the Caucasian singleton and assessment of fetal growth: an analysis of a total population perinatal database. BMC Pediatr 2005; 5(1):13.

(448) Mange KC, Feldman HI, Joffe MM, Fa K, Bloom RD. Blood pressure and the survival of renal allografts from living donors. J Am Soc Nephrol 2004; 15(1):187-193.

(449) Ataei N, Agamohammadi A, Yousefi E, Hasseini M, Nourijelyani K, Tayebi M et al. Blood pressure nongrams for school children in Iran. Pediatr Nephrol 2004; 19(2):164-168.

(450) Stein AD, Barnhart HX, Wang M, Hoshen MB, Ologoudou K, Ranokrishnan U et al. Comparison of linear growth patterns in the first three years of life across two generations in Guatemala. Pediatrics 2004; 113(3 Pt 1):e270-e275.

(451) Bonsu BA, Harper MB. Differentiating acute bacterial meningitis among children with cerebospinal fluid pleocytosis: a multivariable regression model. Pediatr Infect Dis J 2004; 23(6):511-517.

(452) Healey C, Osler TM, Rogers FB, Healey MA, Glance LG, Kilgo PD et al. Improving the Glasgow Coma Scale: motor score alone is a better predictor. J Trauma 2004; 54(4):671-678.

(453) Odibo AO, Marchiano D, Quinones JN, Riesch D, Egan JF, Macones GA. Mild pyelectasis: evaluating the relationship between gestational age and renal pelvic anterior-posterior diameter. Prenat Diagn 2003; 23(10):824-827.

(454) McDonald SP, Collins JF, Johnson DW. Obesity is associated wiht worse peritoneal dialysis outcome in the Australia and New Zealand patient populations. J Am Soc Nephrol 2003; 14(11):2894-2901.

(455) Chitty LS, Altman DG. Charts of fetal size: kidney and renal pelvis measurement. Prenat Diagn 2003; 23(11):891-897.

(456) Wolfe F, Pincus T. The level of inflammation in rheumatoid arthritis is determind early and remains stable over the longterm course of the illness. J Rheumatol 2001; 28(8):1817-1824.

(457) Stone RA, Youk AO, Marsh GM, Buchanich JM, McHenry MB, Smith TJ. Historical cohort study of US man-made vitreous fiber production workers: IV Quantitative exposure-response analysis of the nester case-control study of respiratory system cancer. J Occup Environ Med 2001; 43(9):779-792.

(458) Pijoan-Zubizarreta JI, Irigoien-Garbizu I, Aguirre-Errasti C. Intervalos de referencia poblacional y determinantes de la homocisteina plasmatica [Population reference ranges and determinants of plasma homocysteine levels]. Med Clin (Bar) 2007; 117(13):487-491.

(459) Bland JM, Lees C, Bland JM, Espinoza-Dorado J, Moscoso S. Fetal Doppler velocimetry at high altitude. Ultrasound Obstat Gynecol 2001; 18(4):329-334.

(460) Wolfe F. A reappraisal of HAQ disability in rheumatoid arthritis. Arthritis Rheum 2000; 43(12):2751-2761.

(461) Royston P, Sauerbrei W. Building multivariable regression models with continuous covariates in clinical epidemiology - with an emphasis on fractional polynomials. Methods Inf Med 2005; 44(4):561-571.

(462) Friis-Moller N, Reiss P, El-Sadr W, d'Arminio Monforte A, Thiebaut R, De Wit S et al. Exposure to PI and NNRTI and risk of myocardial infaction: Results from the D:A:D Study. 13th Conference on Retroviruses and Opportunistic Infections, Denver, February 2006 2006.

(463) Palella FJ, Chmiel JS, Moorman AC, Holmberg SD, and the HIV Outpatient Study Investigators. Durability and predictors of success of highly active antiretroviral therapy for ambulatory HIV-infected patients. AIDS 2002; 16:1617-1626.

(464) Collett D. Modelling survival data in medical research. 2nd ed. London, UK: Chapman and Hall; 1995.